

26 November 1981

Threats for British academic research

Research councils in the United Kingdom have recently been spared the axe. Is it now about to fall?

Mrs Margaret Thatcher, the British Prime Minister, was proudly telling the Parliamentary and Scientific Committee a year ago that the British government's spending on research had been "protected" from the round of government economies then coming into effect. Will she be able to make the same boast when the government's budget for 1982-83 has been completed? For the past several months, the research councils have been as much concerned with helping British universities out of trouble as with their own affairs. Now, with financial planning for 1982-83 under way, they are alarmed that, next time, the Treasury's finger will point at them.

The research councils have become vulnerable for several reasons, but chiefly because the British government's economic policies have failed. Neither the government nor its supporters can have imagined that it would be embarking on its fourth year in office (next spring) with yet another deflationary budget. By then, the assumption was, the benefits of financial prudence would be sweeping through the British economy. In practice, however, the government failed to implement the controversial but at least novel monetarist policies to which it was committed with anything like the vigour necessary to give them a sporting chance. So the chances are high that the British Treasury will need further reductions of public expenditure in the spring. The research councils are certain to catch its eye.

Two particular arguments will come up, the first of which is the doctrine of equal misery, thoughtlessly used a year ago as the excuse for reducing public support for British universities by 8.5 per cent over three years. In a memorable display of insouciance, the then Secretary of State for Education and Science told the House of Commons that all public institutions must expect to share the burden of financial sacrifice. He was echoed last Wednesday in the House of Commons by his successor, Sir Keith Joseph. Ironically, if the cost of the enforced redundancies among academics now in prospect falls on the government (and there is no other way in which it can be met), the decisions that have been made will not constitute economies at all. The danger now is that the reduction of support for the universities will be used as an argument for cutting back on the budgets of the research councils. For if the research councils exist in part to support academic research, and if the number of academics is to be reduced, it will be argued that research council budgets can be correspondingly reduced.

That argument is a nonsense which betrays a facile ignorance of the upheaval now under way in British universities. Ostensibly at least, the selective allocation of funds by the University Grants Committee in the summer involved an assessment of individual universities' performance in research. At least in theory, casualties among university departments in the next few years are likely to be heaviest among those with an indifferent record in research. On this assumption, the university departments left substantially intact will be just as much in need of research support in the years ahead as in the past. And if the talents of able academic researchers left stranded in moribund departments are not to be wasted (see *Nature* 19 November, p.198), the research councils will also need funds to help people migrate to centres where promising research can be carried through. What this implies is that the reduction of public subvention for the universities is entirely irrelevant to the needs of the research councils. And any substantial reduction of the councils' capacity

to support academic research will further endanger the health of the British academic research enterprise, already well below par.

Although the budgets of the research councils have been to some extent protected since stagnation set in a decade ago, the protection is far from complete. The cash limits laid down for the present financial year have not adequately compensated for inflation. Meanwhile, the effectiveness of what the research councils can afford to spend in universities is being steadily undermined. With the collapse of the dual support system, by which universities are supposed to provide basic research facilities out of their own funds, the research councils are willy-nilly having to dip into their own pockets for that purpose. At the same time, they are finding that they can no longer fund all the research proposals of high quality that come their way. The Medical Research Council now says so explicitly. But even the pattern of grant applications in the past few years, when the research councils have congratulated themselves that good proposals have been adequately funded, has probably been misleading. Academic research groups, which are not foolish, have for many years been trimming their ambitions to the public knowledge of what the research councils can and will support. The knowledge that proposals that are too ambitious are unlikely to succeed has already contributed to the despondency in the British research community.

Even as things are, the condition of the British academic research enterprise gives the lie to the government's claim that research has hitherto been protected. As the infrastructure is eroded, the research councils are compelled to operate on ever thinner ice. Already they have been forced to postpone plans that would in normal times have been considered vitally important — the Science and Engineering Research Council, for example, has just put off a plan to break new ground in the fashionable field of molecular electronics. The danger now is that the collapse of morale among academic researchers will be accentuated by the real decline of resources to support their work. The result, quite quickly, could be catastrophic. Even as things are, the government could find that the lack of enterprise and innovation that has brought Britain to its present economic plight will persist even when the long recession comes to an end. To think of further reductions of the research councils' budgets is to behave as if the chance of economic recovery has now vanished. If that is what the government believes, it should say so.

Hunger strike's damage

Andrei Sakharov is on hunger strike. Is he right?

Two years of exile in the city of Gork'i would be enough to drive anybody to despair. So it is forgivable that even courageous spirits such as Andrei Sakharov and his wife Yelena have embarked on a hunger strike to persuade the Soviet government to agree that their son's fiancée, Liza Alexeyeva, should be allowed to join her future husband in the United States. One snag, as the Sakharovs must know, is that their isolation in a remote part of the Soviet Union to which foreign visitors are not allowed will blunt the effect of their protest. Gork'i is not, after all, Belfast, where the British government obligingly answered questions from the press about the slow progress towards death of the Irish republicans who fatally starved themselves earlier in the



year. The Soviet authorities, on the other hand, will have no such obligation.

Andrei Sakharov's explanation of his despair is, as always, moving but, on this occasion, strangely unreal. His sense of responsibility derives from having urged his son Alyosha to emigrate when a brief opportunity appeared in 1978, leaving his fiancée behind. (They were later married by proxy in Montana.) But the Soviet authorities are habitually so slow to grant exit visas to citizens wishing to join foreign spouses as to support the belief that they are not willing to do so. The Sakharovs know this, and must also suspect that the sacrifice on which they have embarked may prove fruitless: if they should die, Liza Alexeyeva will be even less well-placed than at present. The Sakharovs' courage is admirable and beyond dispute. Is it possible that on this occasion their isolation has led them to misjudge the future?

Andrei Sakharov's explanation lends some support to that unhappy guess. He complains that appeals for help with his immediate problem addressed to four members of the Soviet Academy have come to nothing, and that one of those asked — Academician Ya. Z'eldovich — had flatly refused to help on the grounds of the "shakiness of his own position". Sakharov goes on to accuse Z'eldovich of abandoning responsibility. It is, however, well-known in the West that Z'eldovich, a distinguished theoretical physicist turned cosmologist, is among that large company of Soviet scientists unable to accept invitations to conferences abroad. Sakharov's own protest against illiberality has hitherto been entirely admirable. In the loneliness of his exile, it is natural that he should be looking to others for help. What they do, and how they do it, is nevertheless for them to decide. That the Sakharovs themselves have fallen back on the hunger strike is especially distressing, not merely because of the risks involved but because the hunger strike is essentially an irrational form of protest. Governments may yield on particular points, but are unlikely to recast their policies as a result. More often, they dig in their heels, declining to give in to emotional blackmail.

Academic cat and mouse

Can Sir Keith Joseph, the least successful minister in the British Government, be trusted with the universities?

The proposition that higher education is too important to be left to educationists would be widely accepted. But can the future of British higher education safely be entrusted to Sir Keith



Joseph, the present Secretary of State for Education and Science? In a debate last week in the House of Commons, Sir Keith uttered this perplexing statement:

Those who maintain that it would be educationally less harmful but still productive of the same savings to go slower have not made out their case. I concede that logically such a case could be made, but I do not believe that it has been made.

If one is made I shall consider it.

If words, even as interpreted by *Hansard* shorthand writers, mean anything, he was saying that British universities have not yet found the forms of words with which to defend themselves against the folly of his government's decision that the university budget should be reduced by 15 per cent in three years. At the same time, however, Sir Keith implied that if only academics were as smart as he is, they would get their act together and persuade him, as he secretly knows they could, that their case for a less draconian contraction makes economic as well as academic sense. And secure as he is in the knowledge that no British academic will pass that stringent test, he will let the system founder.

Sir Keith has a curious reputation, more that of an intellectual than of a politician. His spell as Secretary of State for Industry is memorable chiefly for his advocacy of the importance of market forces and for his contradictory willingness to make out a cheque for whichever nationalized industry was then in trouble. Last week, he rightly challenged his critics to explain how an open-ended commitment by British governments to meet the cost of educating all qualified entrants could be reconciled with the limited capacity of taxpayers to meet the cost. The answer (which follows) is not arithmetical.

Sir Keith was right last week implicitly to say that no British government could permanently agree that all "qualified" students should be educated free, in the disciplines of their choice, at the taxpayers' expense. He should also have done the decent thing and said that Robbins, as a principle, is dead. What bothers the government about the open-endedness of its inherited commitment, however, is not so much the cost of operating the universities as that of paying maintenance grants to students. Sir Keith's officials are already negotiating with the National Union of Students on the students' demand that their maintenance grants should be increased by nearly 18 per cent. The officials will have at least one hand tied behind their backs by the legislation that compels each local authority to pay maintenance grants to students whenever they are awarded places at institutions of higher education. The profligacy of this legislation is, outside the United Kingdom, beyond belief. Not merely does it constitute Sir Keith's chief problem but it is the albatross hung around the neck of the university system. Replacing the present system of mandatory student grants, limitless in cost, with a system of scholarships awarded to deserving people would help not merely to solve the universities' problems but would get Sir Keith Joseph off the hook on which he has quickly (but characteristically) impaled himself. Unhappily, neither the universities nor the minister supposedly responsible for their welfare have the stomach for such a fight.

Monument for a giant

Sir Hans Krebs died this weekend in Oxford.

It is quite true that in the 1930s, *Nature* declined to publish Sir Hans Krebs's article on what most people have since called the Krebs cycle. The explanation given was that the article should be placed in a more specialized journal. Krebs himself cherished the rejection slip but modestly and confusingly always referred to the biochemical cycle he first described by its original name, the "citric acid" cycle (see *Nature* 291, 381; 1981).

Krebs, a refugee from Nazi Germany, carried with him Warburg's way of regarding the biochemical problems of metabolism and, perhaps more important, Warburg's techniques. But only Krebs's flair can account for his unmasking of the intricacies of the most rudimentary of the metabolic processes in animal tissues — the oxidation of sugars to form ATP — with the techniques of half a century ago. Inevitably, that single piece of work will be his chief monument. Another is in danger of being overlooked. Krebs, naturally a magnet for students, modestly bent his energies to teach them all he knew. At one stage, more than a dozen British chairs of biochemistry were occupied by his students. He was a giant eager that his shoulders should be used by younger people.

MIT agrees to accept Whitehead grant

Faculty votes yes, but some strings remain

Washington

Despite strong objections from several members of its science departments, the full faculty of Massachusetts Institute of Technology (MIT) voted by an eight-to-one margin last week to approve MIT's acceptance of an offer from multi-millionaire Edwin (Jack) Whitehead to set up an independent biomedical research centre associated with the institute.

The faculty's approval gives a green light to the MIT Corporation and the institute's president, Dr Paul E. Gray, to accept Mr Whitehead's offer of an initial \$20 million for the capital costs involved in building the research centre, and an eventual endowment of \$100 million to cover operating costs. The centre's main goal will be to pursue research into the applications of molecular genetics to developmental biology.

The money will allow the centre, which will be headed by Nobel laureate and MIT biology professor David Baltimore, to recruit at least 20 full-time professors. All will be members of MIT's biology department, since although funding for the centre will be controlled by members of Mr Whitehead's immediate family and other trustees nominated by him, the centre will for academic purposes be treated as an integral part of MIT.

When details of Mr Whitehead's original offer became known early in the summer, several faculty members complained that research staff at the centre would be subject to a "dual allegiance" that could lead to tensions between them and other MIT academics (*Nature* 1 October, p.329). Concern was also expressed at the original proposal made by Mr Whitehead — co-founder and president of the laboratory instrument company Technicon, which was bought by Revlon in August 1980 in a deal that involved the transfer of \$300 million in Revlon stock to him — that the research centre should own all the patents originating from its work and benefit directly from their licensing.

In the intervening period, the terms of the proposed arrangement have been modified to give MIT more direct control of the centre. Thus all new faculty members appointed to the centre will be required to adhere to existing MIT rules on salaries and benefits, and fulfil the conventional teaching and committee responsibilities of other MIT academics.

Mr Whitehead has insisted, however, that the funds for the centre, to be known

as the Whitehead Institute for Biomedical Research, remain under the control of its finance committee and that his three children should remain a majority on the committee. The conventional arrangement is for a benefactor to provide funds directly to a university, but he says that if he had wanted to give the money to MIT and let the institute decide how to spend it, he would have done so.

Not all the critics have been satisfied with the concessions that Mr Whitehead has agreed to. In a letter circulated before last week's meeting, 33 faculty members expressed their continuing "deep concern" about the implications of the proposed arrangements. The letter, drawn up by physics professor Anthony French and biology professor John Buchana complained that the centre's staff would be selected primarily for their potential contribution to its research activities rather than in order to meet MIT's educational needs; that the creation of the centre would lead to an imbalance in the biology department, since it would provide one-third of the full professors; and that there would be a lack of symmetry between the voting power of the centre's scientists over the work of the rest of the biology department and vice versa.

MIT officials, however, are keen to accept Mr Whitehead's offer, which they see as helping to keep MIT in the forefront of biological research and its potential

commercial application. In a letter circulated to all faculty members before last week's meeting, Dr Gray and MIT provost Francis Low said they shared the general concern about the potential problems that could arise as a result of the dual loyalty of faculty members. But they added: "Although this risk cannot be eliminated, in our view the worries are based on worst-case scenarios that are not likely to materialize".

After a sometimes stormy debate, in which Dr Gray said he did not know the detailed reasons why a similar offer from Mr Whitehead had been turned down by faculty members of Duke University in North Carolina in 1974, a motion to support the creation of the research centre was approved on a straw vote by a majority estimated as eight-to-one. At the same time, the motion said that members of the faculty "acknowledge the existence of legitimate, deep concern over the risks inherent in the venture, and hope that efforts to minimize these risks will continue".

Mr Whitehead's offer will now be formally considered by the members of the corporation, the controlling body of MIT, when they meet on 4 December, and is expected to be accepted. Dr Baltimore expects the new institute to get under way as soon as the affiliation had been formally approved by the corporation.

David Dickson

Mitterrand embraces information policy

M. Francois Mitterrand, President of France, is wasting no time responding to the "information shock" — the complex of issues surrounding microelectronics and unemployment.

In June, Mitterrand asked the publisher Jean-Jacques Servan-Schreiber (who had written a book on the subject) to gather a few experts to study the question. Servan-Schreiber's response came last week in a report which proposed the setting up of a world centre in Paris. This would gather together international experts in informatics, and have a budget of around £10 million a year (£4 million of which would pay the 60 staff). Within the week, Mitterrand had agreed and ordered that the details should be worked out within a fortnight. So the World Centre on Informatics should be in action in 1982.

What it will do, however, is not clear. Servan-Schreiber's vision is of a transformation of the world mediated by the development of the personal computer (something cheaper and more accessible than the present office and household machines). Such computers could make work more varied and creative, and provide everyone — including the Third World — with almost unlimited access to information. Development is a matter of

education, he says, and the personal computer can facilitate it. The world centre, therefore, should observe and assist this transformation.

Mitterrand, however, is aware that in industry the application of microelectronics might cause unemployment, and that the computer can interfere with liberty and privacy. At the same time, he considers that the economic development of France, and of other Western nations may depend on the rapid implementation of microelectronics and informatics. Elected to conquer unemployment, and determined to increase his country's economic strength, he needs an answer to this dilemma and hopes that the world centre, as a kind of full-time standing committee of academics, will provide it.

Although the precise form of the centre has yet to be determined, the staff is certain to be international. So far, ten experts have expressed interest in joining — seven from America, two from Scandinavia and one from China.

If the centre develops as Mitterrand wishes, it will:

- Monitor the applications and effects of microelectronics worldwide.
- Forecast the possible role of France in these developments.

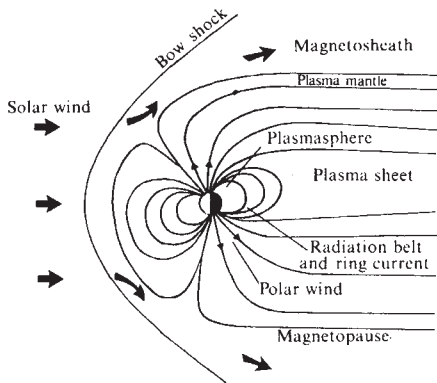
- Study how to educate users and potential users of microelectronics, particularly in relation to maintaining employment.
- Help develop new programs and languages appropriate to a personal computer (Servan-Schreiber version).
- Propose the means of transferring the technology to the Third World.

Robert Walgate

Magnetospheric experiment

UK joins in

The Science and Engineering Research Council has decided to fund participation by the United Kingdom, to the tune of £2 million, in the Active Magnetospheric Particle Tracer Explorers project (AMPTE). The council's Rutherford Appleton Laboratory, in partnership with the Mullard Space Science Laboratory of University College London and other university groups, will be developing the satellite as a separate component of the Ion Release Module satellite being developed by the Max Planck Institutes of



Astrophysics and Aeronomy in West Germany. The two satellites, together with the Charge Composition Explorer satellite being developed in the United States, will all be launched in mid-1984 on a Thor Delta rocket, hastily acquired following space shuttle delays.

The magnetosphere is the region where the magnetic field of the Earth interacts with the charged particles and magnetic fields of the solar wind, the continuous flow of plasma from the Sun. The aims of the experiment are to release lithium and barium ions into the Earth's magnetosphere and, by monitoring their subsequent behaviour, to investigate several fundamental aspects of the magnetosphere as well as to simulate conditions near other planets. The Ion Release Module will be placed in a highly eccentric orbit, so that ions may be released both inside and outside the magnetosphere.

One of the problems to be tackled by the Charge Composition Explorer is to determine whether or not the magnetosphere is "open" — that is, whether the Earth's magnetic field lines can connect with those of the interplanetary medium so that solar plasma can enter the Earth's field

from downwind. To this end, ions will be released just upstream of the bow-shock. The Charge Composition Explorer, in an equatorial orbit, will detect those ions only if the magnetosphere is "open".

The UK satellite will be placed in an orbit with that of the release module, maintaining a separation of 100 km or so. The particle and field detectors on board will monitor the behaviour of the ejected ions immediately after release. Lithium ions, being light, should pick up a large component of the velocity of the solar wind unless disturbed by plasma instabilities.

The release of the barium ions should produce rather more spectacular results. Because they are more massive than most components of the solar wind, the surrounding plasma will be impeded and a glowing comet-like structure will be produced. If the barium ions are released within the high-velocity plasma of the magnetosheath near dawn or dusk, the cloud will be visible from the ground, and will simulate conditions to be found in comets and in the magnetosphere of Venus. When released "downwind" in the magnetotail, where plasma flows more slowly, conditions will be more similar to those near Io and Titan within the magnetospheres of their respective planets Jupiter and Saturn.

Philip Campbell

NASA budget

Small is necessary

Washington

As the National Aeronautics and Space Administration (NASA) locks horns with the Reagan Administration over just how large a budget cut it must face next year, US space scientists are trying to devise a strategy for exploring the Solar System within the tight budget constraints likely to dominate for the next few years.

For the fiscal year beginning 1 October 1982, the Office of Management and Budget (OMB) has told NASA to prepare for cuts of about \$1,000 million in a budget previously projected at about \$7,000 million. Agency officials are strongly resisting this request, arguing that such a cut would severely damage all NASA's programmes; they have already gone over OMB's head to the White House in an attempt to get the proposed cuts reduced before the budget is formally submitted to Congress in January.

However, the prospects for NASA in general — and space science in particular — are likely to be bleak for the next few years. In this context, the agency has set up a Solar System Exploration Committee to take a close look at the type of projects it should be planning for the future. "If something disastrous were to occur to NASA's budget next year, it would be even more important to plan how we should recover", Dr Noel Hinners, the committee's chairman and director of the National Air and Space Museum in

Washington said last week.

The main aim of the committee is to increase the efficiency of planetary and other Solar System missions while reducing the costs, in particular by breaking down the scientific goals of the missions into "smaller bites" than those represented by, for example, Voyager.

One way of achieving this goal would be to build directly on the experience of previous missions. For example, the Ames Research Laboratory is looking at ways of developing Pioneer-type orbiters to some of the inner planets, using derivatives of terrestrial orbiters to study the geological characteristics of the Moon or Mars.

Pioneer-type spacecraft would only be able to achieve relatively limited scientific objectives, and would produce a relatively low data rate. But a mission could be launched for between \$100 and \$150 million, the same order of magnitude as the Explorer satellites currently being successfully used by NASA.

As well as looking at ways of using smaller spacecraft, the committee is investigating how to save money by increasing operations efficiency, for example by reducing the length of trips, compressing the scientific data to be transmitted back to Earth or the sharing of individual pieces of equipment between different research groups.

Four subcommittees will meet next month at the Jet Propulsion Laboratory to work out the type of programme that could be developed on the basis of such components. The subcommittees expect to report to NASA with specific mission proposals by next summer.

Dr Hinners said last week that he had discussed this new approach in some detail with officials at OMB, who expressed general approval but raised two possible objections. First that a list of the various missions which could be included in an overall envelope of between \$300 and \$400 million a year — only slightly above the current level — looked like an ambitious shopping list; and second, that breaking down planetary missions into smaller projects might lack the public appeal of Apollo or Voyager.

To the first objection, NASA replies that it is the size of the effort and the balance between costs and return that should be measured rather than the number of space vehicles flown. The second objection raises a broader question faced by NASA: can the apparently-increasing enthusiasm among the public for space exploration, particularly in the light of the excitement generated by photographs sent back by the Voyager spacecraft, be interpreted as a mandate to increase spending on space science?

Those who do believe that the public will be eager to support ambitious plans for space exploration are still pushing hard for what they consider should be the next major mission, a manned sample-return space vehicle to Mars. NASA administrator James Beggs told the advisory

board last week that he shared the enthusiasm for such a project, although arguing that it should be planned in conjunction with what is likely to be NASA's next major project, a permanent orbiting space station which could be used as a base for manned planetary missions.

Members of the advisory board wondered whether there might be a danger that if NASA did embark on such a project it might further squeeze the space science budget, as the Apollo Moon landings and the space shuttle appeared to have done, but Dr Beggs tried to calm such fears. Admitting that the near-term looked relatively bleak, with no new planetary starts likely to be approved for 1983 or 1984, he replied that in the past the space science budget had done best precisely at those times when NASA had been able to generate political support for its major undertakings.

David Dickson

Diablo Canyon reactor

Licence revoked

Washington

The US nuclear industry has come in for an unexpected roasting from the new chairman of the Nuclear Regulatory Commission (NRC), Dr Nunzio Palladino. His criticism of its quality control was made before a congressional committee on the same day that the NRC voted to revoke a licence issued only two months ago to permit start-up tests at the Diablo Canyon nuclear reactor in California, on the basis of a series of engineering mistakes made when the plant was strengthened to resist earthquakes in the mid-1970s.

Six weeks of frequently-violent demonstrations by anti-nuclear protesters in September and October failed to stop the plant's owners, Pacific Gas and Electric

Company, from proceeding with plans to start to load the reactor with uranium. The protesters claim that the plant is inherently unsafe because it has been built less than three miles from the Hosgri Fault, a branch of the San Andreas Fault system.

As the demonstrations were coming to an end, however, an engineer with the company discovered that a blueprint had been misread when the plant's support structure was being strengthened to compensate for the possibility of earthquake damage, causing the loads on various components to be miscalculated. Further errors found later included the misapplication of stress level numbers along the Hosgri Fault and the use of incorrect data in calculating the ability of pipes to withstand an earthquake.

Besides causing considerable embarrassment to the Pacific Gas and Electric Company, this discovery led to immediate pressure on NRC from Congress. The Reagan Administration has promised to speed up the licensing of new plants; but this strategy requires that public confidence should be maintained that safety is not being sacrificed in the process.

Appearing before the House Interior subcommittee, Dr Palladino, former dean of engineering at Pennsylvania State University, said that "after reviewing both industry and NRC past performance in quality assurance, I readily acknowledge that neither have been as effective as they should have been in view of the relatively large number of construction-related deficiencies that have come to light".

An order approved by the five-man commission, which voted 4-1 to suspend the low-power operating licence and unanimously for an independent audit of the quality control safeguards used by Pacific Gas and Electric Company, said that the suspension was necessitated by the seriousness of the errors in the initial review of safety modifications.

Following NRC's decision, the Pacific Gas and Electric Company issued a statement saying that it was "disappointed", since "nothing has been discovered to date that would indicate that the plant is not safe". The company claims that the plant had many redundant safety systems compensate for any threat from the Hosgri Fault.

David Dickson

More bickering about solar satellites

Washington

The National Aeronautics and Space Administration (NASA) seems unlikely to grant the request from the European Space Agency (ESA) to bring forward the launch of the International Solar Polar Mission, now due in 1986.

ESA had made this request following NASA's decision to stop work on the vehicle which it was to have provided for a dual-spacecraft mission, with the original intention that the two would pass simultaneously in opposing orbits over the poles of the Sun. European scientists had hoped that NASA, which will still launch the ESA space vehicle from the space shuttle, would be able to arrange an earlier flight to help compensate for the disruption and the loss of experiments which the cancellation has caused.

Last week, however, NASA officials said that the growing concern over whether the shuttle will be able to maintain an already overcrowded launch schedule means that a 1984 launch is virtually out of the question.

There is a slight possibility of a launch in 1985. However, since there is only a relatively short launch window in that year, and since the same window is required by the Galileo probe and orbiter scheduled to start its journey to Jupiter at the same time, the chances of arranging both launches within the same period seem slim.

ESA officials are still angry at the way in which NASA cancelled its proposed spacecraft; a resolution adopted by ESA's space programme committee last month suggested that ways should be explored of seeking some type of compensation from the US agency and ESA's secretary general Dr Quistgaard suggested the early launch date as one form of recompense.

US officials admit that it is the first

time NASA has had to go back on a previous agreement, but claim that the "memorandum of understanding" signed between the two organizations makes it clear that the solar mission agreement was subject to the availability of funding, and that NASA could not legally commit itself to more than one year's advance funding.

In addition, Dr Hans Mark, deputy administrator of NASA, told a recent meeting at the National Academy of Sciences that although the decision to cancel NASA's involvement in the solar mission was regrettable, there had been several occasions in the past when European governments had broken commitments, for example some of those made through the North Atlantic Treaty Organization.

On the first point, ESA claims that the memorandum of understanding had been agreed on the basis that normal funding procedures would be pursued by both sides — and that NASA is at fault not for having failed to provide the funding promised, but for having decided unilaterally to omit the project from its request to the Office of Management and Budget for the 1983 budget. On the point that some European organizations have broken commitments in the past, ESA argues that it is not fair to penalize ESA for failures of individual European states.

Meanwhile NASA scientists are discussing the possibility of launching an Explorer satellite into an elliptic orbit around the Sun to coincide with the European spacecraft's encounter. This satellite, which would be launched in 1985, would provide baseline measurements of the solar wind — and that might go some way to making up for the deficiencies resulting from the decision not to send a full US spacecraft.

David Dickson

Uranium enrichment

US-India stand-off

New Delhi

India is trying to become self-sufficient in the production of fuel for its nuclear power plants following its bitter experience with the United States over the supply of low enriched uranium for the Tarapur plant in Maharashtra state.

The latest round of discussions between Indian and US officials has failed to break the impasse over the clearance of at least pending applications with the United

States for the supply of enriched uranium. The United States maintains that India should sign the Nuclear Non-Proliferation Treaty and open all its nuclear installations — indigenous as well as foreign-aided — for international inspection as required by the US Nuclear Non-Proliferation Act of 1978. The supposed fear is that a uranium reprocessing facility in India might be used to extract plutonium for atomic weapons.

India rejects this contention, however, arguing that the 1978 US legislation should not be applied retrospectively and unilaterally to a bilateral agreement entered into in 1963. India has said time and time again that its nuclear technology would be used for peaceful purposes only. India holds the Nuclear Non-Proliferation Treaty to be discriminatory, saying it includes only civilian establishments and specifically excludes military establishments of the nuclear weapon states which prescribe non-proliferation for others and not for themselves.

The issue is now a matter of principle — especially as India is now almost self-reliant for nuclear fuel production.

Indian nuclear scientists have developed mixed oxide fuel of uranium and plutonium which can work as alternative fuel in place of the enriched uranium supplied by the United States for the Tarapur plant. The only other operational nuclear power plant at Kota in Rajasthan utilizes indigenous natural uranium. The nuclear plants being built at Narora and Kalpakkam will also be pressurized heavy water reactors using indigenous uranium.

Sunil Saraf

US nuclear technology

Exports raise fears

Washington

Fears are mounting in Washington that the Administration's efforts to increase nuclear technology exports could be encouraging the proliferation of nuclear weapons. Last Thursday, members of Congress questioned the Administration closely about its agreement with Australia which, for the first time, would mean the United States sharing its knowledge of centrifuge technology for enriching uranium.

The criticism came only a few days after a new storm had broken over the ability of the International Atomic Energy Agency (IAEA) in Vienna to provide satisfactory safeguards against the diversion of nuclear materials from civilian to military use.

The decision to share enrichment technology with Australia is part of an effort to encourage US companies to participate in a joint venture with the Australian government to construct enrichment facilities for its nuclear industry. It was contained in a memorandum signed on 12 November by President Reagan which also instructed the Department of Energy to look at ways of

British academics at the barricades

Genteel academic militancy reached boiling point last week, with a mass lobby of the British Parliament by some 10,000 university teachers protesting not merely at the British government's decision that the university budget should be cut but at the uncertainty that remains about the arrangements that may (or may not) be made to deal with redundancies among academics. Some of the participants (see picture) wore fancy dress.

The lobby (on Wednesday, 18 November) coincided with a debate in the House of Commons on the planned reduction of the public subvention for universities, called by the Labour



opposition. One government speaker complained that it would have been more convenient if the debate had been arranged for the following day, so that those inclined to do so would have had a chance to listen to what the lobbyists were saying.

Both occasions followed by a lunch-time break the first appearance of Sir Keith Joseph, the new (since last month) Secretary of State for Education and Science, before the Select Committee on Education, at which he and his retinue of civil servants were unable to put into words a definition of the "Robbins principle", the doctrine that qualified candidates for university entry should be catered for. At the beginning of last week, the UK Committee of Vice-Chancellors also (unusually) made public its own account of an unsatisfactory meeting with the minister and a waspish letter it had written to him afterwards.

The debate in the House of Commons has confused and not clarified the immediate financial prospects of British universities. Sir Keith Joseph and his minister with special responsibility for higher education, Mr William Waldegrave, declined to answer the apparently simple question whether the government would pay the cost to universities of breaking contracts with tenured academics. Each of them said, however, that the British government would be prepared to "listen to" arguments that it would save money by extending the period over which the universities were now required to contract.

The Committee of Vice-Chancellors is now drafting such a document.

transferring the federal uranium enrichment programme into private hands.

During a hearing of the Senate Foreign Relations Committee's subcommittee on energy and nuclear non-proliferation, several members questioned Administration officials closely on this decision. Centrifuge technology has previously been subject to strong government restrictions, on the grounds that it could provide a relatively inexpensive way of producing weapons-grade nuclear fuel.

However, the Administration continues to insist that, although a hard line will be taken with any country that diverts civilian technology to military use, in general IAEA provides the best way of minimizing the risks of proliferation through its safeguards and regular inspections.

This argument suffered a setback earlier this year when an ex-IAEA inspector, Mr Roger Richter, told the same Senate committee that IAEA had failed to detect

efforts by the Iraqi government to work clandestinely on nuclear weapons, and that present IAEA safeguards were "totally incapable of detecting the production of plutonium in large-size material test reactors".

At the time, IAEA officials fiercely contested Mr Richter's conclusions, claiming that he had not been aware of all the relevant facts. However, it now looks as if they will have to go through the same process in defending themselves against criticisms made by another ex-inspector, Mr Emanuel R. Morgan, in a report commissioned for the Nuclear Regulatory Commission by commissioner Mr Victor Gilinsky.

The report — not officially released but leaked to the *New York Times* — echoes Mr Richter's conclusion that IAEA is incapable of detecting the diversion of a significant quantity of nuclear fuel "in any state with a moderate to large nuclear energy establishment".

This memorandum, likely to be discussed shortly in a congressional hearing, is the first critical assessment of IAEA's safeguards to have been prepared for a federal agency. The State Department issued a statement saying that although it accepted that the safeguards system was not perfect, there was "simply no alternative to an international safeguards regime".

David Dickson

Product development

British battle on

Brussels

The British government is fighting hard to exclude development risks from the EEC's draft directive on product liability now being debated in Coreper — the Committee of Permanent Representatives to the European Community. This has caused great consternation among the European consumer lobby group, Beuc. The consumer group fears that the exclusion of development risks would undermine the whole system of direct liability, by reversing the burden of proof and placing it on the consumer.

The alternative to development risks, the state-of-the-art defence, means that if the manufacturer can prove that he took all reasonable care in the light of the state of scientific and technical development to ensure that his product was safe, he avoids paying compensation. But who then compensates the consumer for his loss if a product, such as thalidomide, turns out to be dangerous? In the United Kingdom the Pearson Royal Commission and the Law Society, concluded that the manufacturer should bear direct liability for development risks. The Council of Europe has come to the same conclusion and France, Belgium and Luxembourg have signed a convention to that effect.

Belgium takes over the presidency of the Council of Ministers from Britain next year and the United Kingdom is anxious to have this part of the directive dealt with by then. The British argue, following pressure from the pharmaceutical industry, that development risks inhibit innovation and competitiveness. In the last debate on this subject in the House of Commons, Sally Oppenheim, the Minister for Consumer Affairs, argued that the EEC proposals would involve a major change for British manufacturers.

The experience in other EEC countries shows that the change is not as painful as alleged. For instance, the German law on pharmaceutical products provides for direct liability and manufacturers cover themselves by taking out an insurance amounting to 2 per cent of turnover.

The British are, however, confident that a compromise can be reached. This could involve excluding development risks or limiting them to pharmaceutical products.

Jasper Becker

US solar energy

Golden housing

Washington

Denying charges that the Reagan Administration is ideologically biased against solar energy in favour of nuclear, US Energy Secretary Mr James Edwards last week announced the go-ahead for the Solar Energy Research Institute (SERI) in Golden, Colorado, to start building a permanent research and test facility.

The Administration's decision is its first bit of good news for supporters of solar energy, whose budget within the Department of Energy has dropped from the \$707 million proposed by President Carter to less than half that in the current year.

Much of this reduction in funds had been felt at SERI. This institute, which is run for the Department of Energy by the Midwest Research Institute, has already had its budget cut from \$120 million to \$50 million, and staff reduced from 960 to 650.

Since its creation in 1977, SERI has occupied temporary premises in the town of Golden, just outside Denver, Colorado. But it has also been planning a combined laboratory and showcase, to be located on a nearby mesa on land donated by the state.

Initial proposals for a solar research and demonstration centre costing \$124 million were reduced by the Carter Administration, which then requested \$24 million in the 1981 budget for the first stage of construction. One of the first acts of the Reagan Administration was to withhold all construction money on the grounds that it was re-evaluating its whole approach towards funding solar energy research.

At one point, it was rumoured that the Office of Management and Budget was contemplating eliminating all funding for such research from the Department of Energy's budget. Publicly, however, the Administration has stated that although it is eliminating demonstration projects which it feels ought to be supported by the private sector — if at all — it accepts the need for federal support of long-range, high risk research.

The permission from the Department of Energy for construction of SERI's laboratory is being used by the Administration to symbolize a commitment to solar power. But Mr Edwards said that the \$9.5 million which he was making available for the initial construction phase — money from appropriations previously authorized by Congress — did not necessarily mean extra funds for energy research.

But the decision to fund SERI's laboratory does not necessarily mean that federal support for solar energy research will be generous. The rumour in Washington is that the Department of Energy's budget request for 1983 contains a meagre \$93 million for all solar energy work, a contraction to the level of the early 1970s.

David Dickson

Californian Medflies

Innocent victims

Palo Alto, California

If aerial spraying with the insecticide malathion is used to counter the threat of Mediterranean fruit fly in California this winter, an unexpected casualty might be one of the oldest continuously studied insect populations in the world.

Jasper Ridge is a 1,200-acre natural wilderness to the west of the Stanford University campus in California, between suburban San Jose and the city of San Francisco. More than 120 dissertations and papers have been written about the area since 1897 when studies began in earnest. So far, Jasper Ridge has been a virtually unsprayed island, surrounded by a 600-foot buffer zone. But in July spraying began in areas next to the reserve.

Most immediately at risk, of course, are the insects themselves, notably the checkerspot butterfly *Euphydryas editha*, which has been the subject of a continuous population study since 1960, under Professor Paul Ehrlich.



Paul Ehrlich may search in vain next year

Although Jasper Ridge has not been sprayed directly, minute droplets of malathion bait (20–40 μm) have been found on plant leaves within the reserve, and these droplets must have drifted into the area from nearby spraying operations. During the summer, the drifting droplets did not appear to affect the checkerspot butterflies, which lay dormant under ground. But in the rainy season the checkerspots emerge to feed, and any direct spraying during the winter would almost certainly mean their complete disappearance from Jasper Ridge. Aerial spraying has been curtailed for the winter months, but will be stepped up again next spring (see *Nature* 12 November, p.103).

Direct spraying would almost certainly be carried out, however, if two adult Medflies, or a pupa, were found within the Jasper Ridge reserve. The ecological balance of the reserve would then be irretrievably disturbed by an uncontrollable variable. Donald Kennedy, a biologist and Stanford University president, compares the possible loss of Jasper Ridge with the loss of an irreplaceable library. **Charlotte K. Beyers**

CORRESPONDENCE

Alchemy in cancer

SIR — Your editorial (*Nature* 5 November, p.1) describing the attitude of the sponsors of cancer research in the United States towards those who perform the research was most timely. The relationship between sponsor and sponsored in research generally is presumably similar except that in cancer research the goal is more visible and unease at the failure to achieve it more acute.

Without wishing to push the analogy too far, there is an apparent similarity in the relationship between workers in cancer research and their sponsors on the one hand, and on the other, the alchemists pursuing the philosopher's stone at the behest of their princely masters. It is a sobering thought that many alchemists lost their heads for failing in their sphere of endeavour. Fortunately, nowadays, it is only our jobs that are at risk.

NEVILLE WILLMOTT

Department of Oncology,
University of Glasgow, UK

Creationism

SIR — Our letter "American Creation" (*Nature* 2 July, p.95) appears to have resulted in two misunderstandings. The first, implied by the misleading title, is that it was primarily a critique of American Creationism. It clearly was not. The second, that we argue for a "God of gaps" by resorting to the supernatural to fill the gaps in current scientific knowledge, requires more serious consideration, since we agree with Sidney Fox (*Nature* 6 August, p.490) that such a view is a "copout" scientifically and theologically.

In common with most creationists, we acknowledge that faith is based on revelation and on personal knowledge of the Creator. The only empirical evidence to support the assertion that all nature, however adequately or inadequately explained, is God's handiwork, is the fact that all things exist, and show evidence of design and purpose which reflect the revealed nature of the Creator. Given belief in such a Creator we expect to find evidence consistent with his revelation of himself. No doubt an atheist will likewise expect to find evidence consistent with his philosophical position and is likely to leave discordant data to one side in the belief that future study will show it to be either erroneous or, given further data, assimilable to his present beliefs. In that sense it is indisputable that since scientific knowledge is incomplete, gaps exist in any account of the natural world.

Furthermore we would argue that the primary purpose of man's quest to understand origins is philosophical not scientific, and that in a society that will not tolerate a personal infinite God, the purpose of evolutionary theory is often the justification of atheism. In this context it is critical that scientific evidence be analysed within a completely objective framework, before presuppositions are imposed upon it.

Our position is that we believe that existing evidence does not disprove the existence of a Creator and further that the most probable explanation of his activity as Creator is one which does not accord with current evolutionary theory. In examining the factual evidence available in biology we consider that the correct objective approach to biological

systematics is to employ cladistics, a methodology which might usefully be applied also to protein and nucleic acid sequences. The interpretation of such an analysis is subjective, resting on philosophical presuppositions. So any evolutionary relationships that exist are properly deduced from cladistics.

So far as the origin of life is concerned we suggest that the most likely interpretation of scripture is that life did not evolve from a prebiotic soup. Thus we are not surprised that there are difficulties in proposing plausible chemical mechanisms to support the spontaneous generation position. For example, to take Fox's experimental model, we agree that while plausibility can be argued indefinitely, nonetheless the only basis on which the plausibility of such a model may be assessed, however subjectively, is by an examination of the experimental data. Although we do not think it appropriate here to enter into detailed technical comments, we find Fox's data (*Biosystems* 12, 55) singularly unconvincing.

C.H. DARNBROUGH
J. P. GODDARD
W. S. STEVELY

University of Glasgow, UK

Malaria debated

SIR — I have read the paper by Chapin and Wasserstrom (*Nature* 17 September, p.181) with interest. I am disappointed with the presentation and discussion of the important subject of malaria resurgence and its relationship to agricultural production.

The authors give a garbled account of the very concept of malaria eradication and especially of the causes of the relative failure of this great endeavour. They imply that the main obstacle to the early achievement of the planned goal was resistance of *Anopheles* to insecticides. This is not so, even though the latter phenomenon played a significant part in technical problems that the World Health Organization (WHO) was facing during the late 1960s. The most comprehensive analysis of the multiple causes of the disappointing progress of malaria eradication in some parts of the tropical world was presented by WHO in 1969¹. It pointed out that administrative, social, economic and financial factors were largely responsible for the resurgence of malaria at the end of that decade, especially in India.

There is nothing "ironical" in the fact that agriculture in India and elsewhere expanded in regions where malaria incidence decreased spectacularly. It is precisely because of the improved health conditions, due to the use of residual insecticides against *Anopheles* vectors, that efficient agriculture became possible not only by rich landlords but also by small farmers. The astounding graph showing the positive association between the use of DDT in India and the striking increase of malaria in 1969–77 cannot be accepted in support of the authors' thesis of a causal relationship between the two factors.

An epidemiological analysis of the resurgence of malaria in India, such as the one carried out by Akhtar and Learmonth², showed that the increase in the incidence of malaria in India from 350,000 cases in 1969 to nearly 2 million in 1973 and then to 5 million

in 1975 was due to other conditions with an adverse affect on the standard of anti-malaria operations in India and in other countries of the subcontinent. The military conflict with Pakistan, the sharp fall in the flow of American aid and the temporary food shortages due to bad harvests took place during that period. There were delays in the allocation of foreign exchange for insecticides and drugs and there was some loss of urgency of the malaria control campaign because of its good results and the emphasis on family planning.

The successful 1965–69 phase of malaria eradication in India left uncompleted four large areas from which the consequent resurgence spread. These were the Rann of Kutch in western India, Madhya Pradesh hill forests, Orissa hill forest tracts and the forested areas of Assam. Akhtar and Learmonth² indicated that densely populated areas, with extensive irrigation and high agricultural production, have shown less malaria than other areas during the period 1970–75. On the other hand, in spite of the continuous use of DDT for agricultural needs, the amount of malaria in India decreased sharply between 1977 and 1980, thanks to better implementation of control programmes.

A series of valuable studies by Indian experts^{3,4} and an Audit Report of the Agency for International Development⁵ attach much less importance to the problem of insecticide resistance and could be quoted as an argument against the thesis so eagerly adopted by Chapin and Wasserstrom.

The adverse effects of excessive use of residual insecticides in agriculture have been known and stressed by WHO for well over a decade. It is quite untrue that WHO did not urge all countries in the developing world to decrease as much as possible (without endangering their food and health programmes) the use of residual insecticides and to introduce alternative methods of pest control. The emphasis on this policy was clearly stated in the 16th Report of the WHO Expert Committee on Malaria⁶ and was repeated and emphasized in the 22nd Report of the WHO Expert Committee on Insecticides⁷ and in every other relevant WHO report.

The serious imputation that the Food and Agricultural Organization (FAO) policy is being influenced by large insecticide-producing companies will certainly be answered by those accused of irresponsibility or other more sinister intentions. The authors use with evident relish the impressive term "integrated control" although it is doubtful that they fully understand its practical implications. Integrated pest management means a combination of chemical, biological and environmental methods. The complexity of implementation of these methods is often difficult to fathom by non-specialists, who invoke the term with more heat than light and over-estimate its universal feasibility. Integrated control methods are certainly successful in some areas with important crops but they must be tailored to local conditions. Their use requires not only a constant assessment of the size of the pest population but also carries with it the uncertainty of

Continued on page 388

NEWS AND VIEWS

MARS



From Philip Campbell

THE exploration of Mars by spacecraft began in July 1965, when Mariner 4 passed 12,000 km above its surface and provided the first direct evidence for the existence of craters, a weak magnetic field and a thin atmosphere. Mariners 6 and 7, approaching the planet to within 3,500 km in 1969, indicated that craters were a predominant land form but also revealed the existence of chaotic terrains of irregular ridges and troughs. Carbon dioxide was identified as an important atmospheric constituent, although geologically speaking the planet appeared to be similar to the Moon.

This impression was altered dramatically in 1971 with the arrival of Mariner 9 (the first spacecraft to orbit another planet) and the Soviet Union's Mars 2 and 3 spacecraft. Mars 3 contained a lander that was successfully placed on the martian surface; regrettably, its television camera failed after 20 seconds. Mariner 9 comprehensively mapped the surface and revealed (when dust storms had cleared) not only craters but channels, volcanoes and canyons, all of super-terrestrial proportions. It became obvious that Mars was a more dynamic planet than had been thought.

Since that time, the planet has been visited by six Soviet and American spacecraft, including the most sophisticated and ambitious of all planetary probes — Vikings 1 and 2. The data provided by the latter orbiters and landers since their arrival in 1976 still occupy the minds of planetologists, as evidenced by the continuing series of regular international colloquia on Mars.

The third of these was held at the California Institute of Technology last August; the full proceedings will appear in the *Journal of Geophysical Research* next year. Several speakers were invited by *Nature* to present articles on topics of their choice, concentrating on particular aspects of Mars' solid body and its gaseous envelope, and reflecting to some extent the ground covered at the meeting. The six articles that follow, therefore, are not intended to represent a comprehensive survey of the planet. Nevertheless, it is hoped that they will provide a vivid impression of the radical advance in our understanding over the past few years, and even rekindle some of the excitement that prevailed when the Vikings touched down.

One topic that is barely mentioned here is martian biology. It should be remembered that the Viking landers are sitting on sites where biological activity was thought most likely to be found. But the door on this topic might be said to have closed at the previous colloquium, when it was concluded that the top 10 cm or so of soils at the two lander sites were bereft of organic compounds. A forlorn refrain running through the following articles is the need for further visits to the planet; when such a return materializes, seekers of life will probably want to drill deeper into the surface, where complex molecules may find existence more congenial.

The only future Mars mission known to be under active consideration is the European Space Agency's Mars orbiter, the Kepler mission. If Europe decides to proceed with it, Kepler should be launched towards the end of this decade. Meanwhile, the Viking 1 lander is reactivated at regular intervals to return pictures such as that on the left. But planetologists are trying to salvage what they can from future budgets (as reported in the *News* section). Will these articles mark the end of this period of unprecedented exploration?

Philip Campbell is a member of the editorial staff of Nature.

The geophysics of Mars: whence the Tharsis plateau?

from Sean C. Solomon

THE planet Mars, since the first global views of its surface were sent to Earth by Mariner 9, has been regarded as a key member of the terrestrial planetary family. Intermediate in size between the larger (the Earth and Venus) and the smaller (Mercury and the Moon) bodies, Mars also appears intermediate in the vigour and duration of its tectonic and volcanic activity. Roughly half of its surface is pock-marked with impact craters and basins that formed, by analogy with the Moon, over 3.8×10^9 years ago. One third of the surface consists of younger volcanic plains, here and there capped by prominent volcanoes. Major concentrations of volcanoes occur in two broad provinces marked by high surface elevations, positive gravity anomalies and an extended history of surface faulting. The largest of these two provinces, the Tharsis plateau, provided a major focus for new ideas and widespread discussion at the colloquium.

As a backdrop to the discussion of Tharsis, it is well to remember the vast difference between the level of our knowledge of Mars and that of either the Earth or the Moon. We know the planetary mean density and, with less precision, the moments of inertia for Mars, yet we know little of the nature of the Martian mantle or the nature and size of any high-density core. Whether Mars has an internal magnetic field is debatable and is unlikely to be settled by existing data. We have gravity and topographical data of limited coverage and variable accuracy for Mars; we nonetheless do not know the nature and thickness of the martian crust. A relative chronology for major geological units has been derived from crater density measurements; the true ages of martian surface features are not known. The geometry and some relative age information have been obtained for the principal tectonic features, but we do not know the true ages of major tectonic events or the present level of tectonic activity. We are completely ignorant of martian heat flow and of the abundances and distribution of heat sources within the planet.

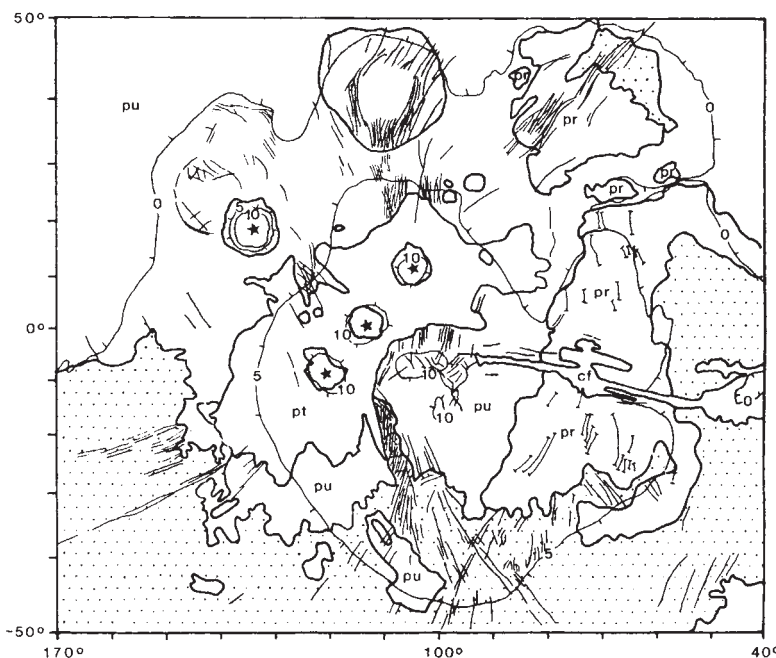
We are therefore faced with the task of understanding the volcanic and tectonic history of Mars in general, and of Tharsis in particular, without the information on the seismic structure or present thermal regime of the interior that we rely on for similar studies of the Earth and Moon. As a further handicap, that history cannot yet be anchored to the radiometric age of a single martian rock sample. The principal

tools at our disposal in this endeavour are images of the planetary surface, gravity anomalies inferred from the acceleration of orbiting spacecraft, and topographical measurements obtained from photogrammetry and Earth-based radar.

The Tharsis region of Mars (see the figure) is a broad plateau 4,000 km in diameter and standing as much as 10 km above the level of the surrounding terrain¹. The plateau is covered by relatively young volcanic plains and includes a number of volcanic constructs large and small. There are numerous extensional fault systems and graben that generally predate the youngest volcanic plains and that display a

of most extensional fractures in and near the Tharsis area.

Recent calculations^{5,6} of the stresses in the martian lithosphere have called the uplift model, as originally proposed, into serious question. Because the Tharsis region is so large relative to the planetary radius, full spherical calculations including both membrane and bending stresses are necessary to model the response of the martian lithosphere at Tharsis to either doming or loading. An uplift model for Tharsis leads to the prediction⁶ of radial compressive features at distances from the plateau centre at which radial graben and extensional features are actually observed.



The Tharsis region of Mars^{1,4}. Contours give topographical height in km above Mars datum. Ancient cratered terrain is indicated by light shading. Major volcanic constructs are indicated by heavier shading; the summits of the four largest volcanic shields, indicated by stars, exceed 25 km in elevation. The volcanic plains units include the ridged plains (pr), the younger Tharsis plains (pt) and the undivided plains (pu). Tectonic features include the extensional faults systems (lines), mare ridges (lines with ticks) and the Valles Marineris canyon system (cf).

crudely radial geometry with respect to the centre of the elevated terrain. There are also tectonic features analogous to lunar mare ridges, presumably the result of horizontal compressive stresses, that are more nearly circumferential to the Tharsis plateau.

The traditional explanation²⁻⁴ for the origin and evolution of the Tharsis region is that broad updoming of the martian lithosphere caused by a thermal or chemical anomaly in the mantle led to fracturing, volcanic emplacement of thin plains units, and finally the formation of the youngest large shield volcanoes. Supporting evidence includes the broad topographical height of the Tharsis region, the large elevation of surface units thought to be relatively old from the density of craters and fractures, and the radial trends

As a result, two alternative models for the origin and evolution of Tharsis have recently been put forward. An 'isostatic' model, first proposed by R. J. Phillips and N. H. Sleep⁷, is based on the hypothesis that the topographical height of the Tharsis plateau, and an additional subsurface excess mass required by gravity data, are compensated isostatically by a low density region of the mantle extending several hundred kilometres deep. The low mantle densities may be partly thermal in origin; A. A. Finnerty and R. J. Phillips also described a mechanism whereby the low densities would result from the depletion of lower melting-point basaltic material in the mantle beneath Tharsis.

A contrasting 'flexural' model for Tharsis, proposed by Head and Solomon⁸, is based on an extrapolation to a larger

scale of the geological processes seen in a number of areas on Mars: volcanic loading of the lithosphere, faulting in regions of thin lithosphere or in response to the flexural stresses resulting from large loads, and a concentration of continued volcanic activity in regions of widespread extensional failure. According to this model, much of the topography of Tharsis is due to volcanic construction, and a portion of the topographical load is currently supported by the finite strength of the martian lithosphere.

How can these two contrasting models be distinguished with our existing information on martian structure and evolution? One useful exercise, conducted by J. B. Plescia and R. S. Saunders⁹ and independently by R. A. DeHon, is to estimate the thickness of the volcanic plains units on the Tharsis plateau from the rim heights of craters partially buried by the plains material. These studies yield estimates of 3–5 km for the maximum thickness of the Tharsis volcanic plains that postdate the heavy impact bombardment of Mars (which probably ended about 3.8×10^9 years ago). Left unanswered by such studies, however, is the question of whether the heavily cratered surface underlying the young volcanic plains in the Tharsis area may itself represent the accumulation of older volcanic units. Thus the total thickness of volcanic material in the Tharsis region is unknown.

The tectonic history of the Tharsis region — the formation of faults of particular types and geometries in particular locations at different stages in martian geological history — holds important clues to the mechanisms responsible for Tharsis. R. S. Saunders and J. B. Plescia described evidence for distinct stages in the development of Tharsis-related graben and extensional faults; each stage was marked by a different centre of activity. T. R. Watters and T. A. Maxwell addressed the question of the relative ages of mare-type ridges and extensional features where the two types of fault system intersect. Although the evidence that one type of fault is consistently older than the other is equivocal at present, the issue is very important for our understanding of the evolution of Tharsis, since the two types of fault cannot be produced at a given location by a single stress field.

A promising line of research is to combine the information on the history of faulting with calculations of lithospheric stresses as constrained by topographical and gravity data. Calculations by R. J. Willemann and D. L. Turcotte and by R. J. Phillips and W. B. Banerdt have shown that models in which the Tharsis plateau acts as a load on the martian lithosphere predict stresses in broad agreement with observed tectonic features. Phillips and colleagues are finding evidence that older fault systems in the Tharsis area fit best with an 'isostatic' model, while younger fault systems fit better with a 'flexural'

model. The need to keep open the possibility that Tharsis tectonic activity may have also been influenced by global-scale processes was underlined by P. H. Schultz and A. B. Lutz-Garihan, who argued that layered deposits near the martian equator once marked the location of the planet's rotational axis; large-scale polar wander on Mars would have led to global stresses in the martian lithosphere sufficiently large to affect tectonic activity in Tharsis and elsewhere.

If there was a consensus at the colloquium on the nature and evolution of the Tharsis region, it was that none of the simple models for Tharsis topography and tectonics (uplift alone, purely isostatic support, or solely volcanic construction and lithospheric flexure) is likely to be completely adequate. Continued work towards refining and testing these alternatives is clearly warranted. Tasks of high priority for martian geophysical studies in the near future include further elucidation of the chronology of tectonic

activity, continued reduction of Viking Orbiter gravity data, new measurements of martian topography with Earth-based radar, and continued development of quantitative stress models to test hypotheses for Tharsis against these new data. For the longer term, our understanding of martian geological evolution will be limited by the large gaps in our information on the planet's surface and interior. Those gaps can be filled only by returning to the planet with new spacecraft experiments. □

1. Scott, D.H. & Carr, M.H. USGS Map I-1083 (1978).
2. Phillips, R.J., Saunders, R.S. & Conel, J.E. *J. geophys. Res.* **78**, 4815 (1973).
3. Carr, M.H. *J. geophys. Res.* **79**, 3943 (1974).
4. Wise, D.U., Golombek, M.P. & McGill, G.E. *Icarus* **38**, 456 (1979).
5. Phillips, R.J. & Lambeck, K. *Rev. Geophys. Space Phys.* **18**, 27 (1980).
6. Turcotte, D.L., Willemann, R.J., Haxby, W.F. & Norberry, J. *J. geophys. Res.* **86**, 3951 (1981).
7. Sleep, D.U. & Phillips, R.J. *Geophys. Res. Lett.* **6**, 803 (1979).
8. Solomon, S.C. & Head, J.W. *Lunar Planet. Sci.* **11**, 1063 (1980).
9. Plescia, J.B. & Saunders, R.S. *Proc. 11th Lunar Planet. Sci. Conf.* 2423 (1980).

Volcanism on Mars

from James W. Head III

SINGLE volcanic eruptions, such as the awesome Mount St Helens event, and the frequent spectacular basaltic eruptions on Hawaii, serve to focus our attention on active volcanic processes. Instructive as these local events are, we should not lose sight of the basic reasons why the study of planetary volcanism is important.

If we were to make a list of fundamental things we wanted to know about a planet, two questions would appear high on the list. First, what is the planet made of, and second, how does the planet generate and get rid of its heat? Complete answers to these two fundamental questions require detailed and direct examination of the interior of a planet and a knowledge of how that interior has changed with time. Since it is extremely difficult to examine the interior of a planet directly, we are forced to rely on largely indirect means to answer these questions. The record of planetary volcanism on the surface of a planet is the most important indirect evidence we have.

For example, the record can tell us a lot about the composition and state of the interior. The mere presence of volcanic deposits implies partial melting in the interior and a state of stress in the lithosphere favourable to eruption. In addition, volcanic deposits supply compositional information directly in the form of samples, and indirectly through remote sensing and the interpretation of volcanic morphology. Information on thermal evolution can also be obtained — we

might, for example, compare the volcanic record of the Moon, where 100 per cent of the surface volcanic deposits were emplaced in the first 50 per cent of history, with the Earth, where 70 per cent of the surface volcanic deposits were formed in the last 5 per cent of Solar System history. Clearly, the two bodies had different thermal histories. The individual deposits on a planet and their variation with time are thus important clues to a planet's thermal evolution. Volcanism can also be related to other processes — volcanism and tectonism are often closely interlinked, as in the dramatic Tharsis region of Mars (see review by Solomon), and the volatiles comprising planetary atmospheres must be tied in some fundamental way to planetary outgassing and the history of volcanism.

These general themes raise specific questions about volcanism on a single planet such as Mars. The first task at hand is the identification and characterization of volcanic deposits. The range of morphological features (for example, shields, cones, flows) must be established, and some quantitative measure of these morphologies and estimates of their relative abundances made. At the conference, R. Pike and G. Clow (US Geological Survey) reported quantitative measurements of martian volcanoes and comparisons with similar structures on Earth, concluding that there is good evidence for three separate classes of volcano on Mars and that most martian shields differ substantially from terrestrial shields. These data are of the utmost importance for the characterization and

James W. Head III is in the Department of Geological Sciences, Brown University, Providence.



Oblique view of the massive martian shield volcano Olympus Mons. Recent topographical data (S. Wu *et al.* 3rd *int. Colloq. on Mars*, 287-289), indicate that the peak rises some 26 km above Mars datum, and that the caldera complex is about 80 km in diameter. A scarp up to 8 km in height surrounds the 600 km wide shield.

understanding of eruption conditions. C.A. Wood (NASA Johnson Space Center) compared calderas on Mars and Earth and noted the relatively low abundance on Mars, attributing it to the lack of both plate tectonic subduction zones and associated abundant silicic volcanism on Mars. S. Wu and co-workers (USGS) presented a topographical map of the massive Olympus Mons shield volcano (see figure), and E. Morris (USGS) described much new photographic data on the structure of this shield, the nature of its basal scarp (rising 4-8 km above the surrounding plain), and on the distribution and characteristics of the thin, low-viscosity flank flows. M. Zeitner (University of Washington) presented evidence that some of the circular craters associated with graben on Mars may represent maar-type eruptions. A study of the distribution and properties of small cones on Mars (H. Frey and M. Jarosewich, NASA Goddard Space Flight Center) provided data to help distinguish between origins as cinder cones, pseudo-craters and pingoes. Variations within and between regions suggest multiple origins for the martian examples. P. Spudis and R. Greeley (Arizona State University) des-

cribed the geology of Tyrrhena Patera, an ancient central vent volcano in the cratered terrain, and interpreted its structure as representing a style of martian pyroclastic volcanism that may have been more common, but not abundant, in early martian history. G. Schaber, K. Tanaka (USGS) and J. Harmon (National Astronomy and Ionosphere Center, Arecibo) reported topographical data from Earth-based radar observations and several new volcanotectonic collapse features and previously unmapped lava flows within Syrtis Major Planitia. The topographical data show a dramatic 5 km rise near the edge of Syrtis Major, rather than the gently sloping surface previously thought to represent the topography. The authors suggest that the elevation of the entire Syrtis Major feature, the previously reported absence of any basin rim mountains or rim ejecta, the rapid decrease in elevation into Isidis Planitia, and the absence of any strong gravity anomaly centered on Syrtis Major, all indicate that the early geological history of this feature does not include the formation of an impact basin as postulated earlier. The feature appears to consist instead of a vast volcanic sequence whose source fissures

(probably concentric to the Isidis Basin), and subsequent extrusive episodes are related to crustal fracturing or fracture rejuvenation following the Isidis impact event early in martian history.

Characterization also requires data on the composition and physical properties of these units. Here remote sensing plays an extremely important role. On the Moon, we have nine sample return sites, yet global characterization of volcanic deposits is still mainly derived from remote sensing data. On Mars we have no sample return sites so we must rely on remote sensing data even more. Remote sensing can provide data on the composition and physical properties of surface materials and can also be used to distinguish, characterize and map surface units. H. Kieffer (USGS) reported a series of global maps of Mars data that have been compiled to similar scales; they will be of great importance in global mapping of volcanic deposits. J. Zimbelman and R. Greeley (Arizona State University) studied the surface properties of the Ascræus Mons shield using remote sensing data and showed that its surface is characterized by a uniform average particle size, arguing against the presence of localized ash deposits. T. McCord (University of Hawaii) and co-workers used Viking approach images to define global surface units and a range of other information — Viking orbital, Earth-based telescopic spectral and albedo data — to characterize these units. The maps emphasize the composition of surface units and differ from traditional geological maps in that they include condensates and eolian deposits as well as bedrock units. The approach holds great promise for distinguishing and characterizing volcanic units on Mars.

Another promising area is the investigation of Earth samples with similar spectral properties to surface units on Mars. D. Evans (University of Washington), R. Singer (University of Hawaii) and co-workers showed that certain weathered basalt tephra and a basalt coating from Hawaiian rocks have similar spectral characteristics to surface units at Viking Lander 1 and to some global units. The final results for the chemical composition of martian fine-grained soil from the X-ray fluorescence spectrometers at the Viking landing sites [B. Clark (Martin Marietta) and A. Baird (Pomona College)] can be interpreted as representing a primary crustal rock type, a mafic flood basalt rich in pyroxene.

An understanding of the range of eruption conditions represented by the deposits and their significance in terms of magma types, volatiles and the state of stress in the lithosphere can be approached by looking for direct morphological analogy with features on Earth. This approach may help to establish the volcanic nature of major landforms, such as the shields, but may be quite misleading for understanding the eruption conditions

leading to specific features on Mars. We know, for example, that gravity and atmospheric density differ between Earth and Mars and can markedly influence eruption conditions and resulting landforms. It is important, therefore, to consider the process of ascent and eruption of magma on Mars from first principles and to develop from these theoretical analyses predictions of the types of volcanic landforms that might be produced in the martian lithospheric and atmospheric environment. L. Wilson (University of Lancaster) and J. Head (Brown University) presented theoretical analyses of the behaviour of both volatile-free magma and magma containing volatiles and the resulting deposits and landforms. A specific example interpreted as a relatively young pyroclastic air-fall deposit on the flanks of Hecates Tholus was described by P. Mouginis-Mark, (Brown University) and Wilson and Head. From deposit characteristics and models of explosive eruptions, possible eruption conditions for this unique deposit were outlined. Although ash deposits had been thought of as candidates for various units on Mars, this appears to be the first identification of a specific deposit from a source region.

Armed with data on the characteristics of deposits and the range of eruption conditions, it would be useful to know the distribution and volume of volcanic deposits in space and time. Greeley and Spudis (*Rev. Geophys. Space Phys.* 19; 13-41, 1981) have recently reviewed the field and it is obvious that we have a long way to go to identify and characterize deposits and to relate them to the impact crater chronology. Progress in this area was reported by P. Francis (Lunar and Planetary Institute, Houston) and C. Wood (NASA Johnson Space Center), who reviewed evidence that silicic pyroclastic volcanism on Mars is unlikely to be a major process.

J. Plescia and S. Saunders (Jet Propulsion Laboratory, Pasadena) and D. Scott and K. Tanaka (USGS) have recently compiled stratigraphies of the volcanic deposits in the Tharsis region of Mars. Although information is accumulating, our understanding of the significance of these data for the thermal evolution of Mars is hampered by a lack of returned samples and of a knowledge of absolute ages. L. Martin (Lowell Observatory) reviewed evidence for possible volcanic activity on Mars during the Viking mission. Despite the excellent geological maps of the USGS, Dave Scott (USGS) showed that many volcanoes and volcanotectonic structures in the western hemisphere of Mars could be mapped using newer Viking Orbiter images, particularly in the older parts of the martian crust.

Where do we stand in terms of achieving the basic objectives outlined above? On the basis of the conference results, it is obvious that much more work of the type reported at the conference is required to characterize

the martian volcanic record. We need to integrate the full range of remote sensing data in order to complement surface morphology with information on composition and physical properties. Some of this information is available, but much remains to be collected from Earth-based observations and from spacecraft near, and on, Mars.

Some of the basic theory of magma ascent and eruption on Mars has been worked out but much work needs to be done to compare present deposits with the theoretical predictions. This will help to establish the full range of eruption conditions and the implications for magma composition and the role of volatiles from both deep and shallow sources. The dis-

tribution of volcanic units in space and time is becoming better known but is far from a level which can constrain thermal evolution. Continued comprehensive mapping is required. A major problem is the nature and level of volcanic activity in early martian history during the time of high impact flux. Even with such information in hand, an absolute chronology is required before the thermal evolution of Mars can be understood.

In summary, progress towards these goals can be made with existing data, but major strides in these areas of fundamental importance to the evolution of planets requires new data of a variety of types, data obtainable by returning to Mars with properly instrumented spacecraft. □

Martian geology

from Michael H. Carr

In addition to volcanism a wide range of geological topics, including channel formation, subsurface migration of water, simulation of martian ejecta patterns and sedimentation at the poles, were discussed at the colloquium. Surficial deposits such as the regolith and the relatively mobile eolian cover received most attention, possibly because a wealth of new experimental and observational data now allows them to be treated in quantitative terms.

Several papers were concerned with weathering. It has been suggested that UV light has an important role in chemical breakdown of materials on the martian surface but new data (M.C. Booth and R.V. Morris, Johnson Space Center) show that under martian conditions, with modest partial pressures of H_2O , oxidation of magnetite, dehydration of goethite and breakdown of silicates take place at rates independent of UV radiation. Other experiments (T.R. Blackburn) show that breakdown of sulphides to sulphate can also occur without irradiation but breakdown is more rapid in samples irradiated by UV. R.L. Huguenin (University of Massachusetts) expanded upon his frost weathering model in which dissociation of adsorbed water on mineral grains and migration of hydrogen ions to lattice defects allow recombination of the OH^- radicals left on the surface to produce H_2O_2 — possibly the active oxidant needed to explain the Viking biology results. The reactions he proposes are all exothermic and do not depend on UV radiation, suggesting the role of UV in martian weathering is less important than previously thought.

The precise nature of the weathering products remains uncertain. Attempts to

match telescopic reflection spectra of Mars bright regions with various terrestrial materials have shown the Mars spectra to be inconsistent with ferric oxides or iron-bearing clays. The best fit is with palagonite, a largely amorphous reaction product of basaltic lava with water which had been previously suggested by several workers, including the Viking Inorganic Analysis Team, to be a possible Mars surface material. However, A. Bonin reported work at the Hebrew University, Israel, on simulation of the results of the Viking biology experiment. He showed that Fe-rich clays can catalyse decarboxylation and/or oxidation of organic acids and was able to reproduce accurately the Viking labelled release results. Fe-rich clays were also used successfully to simulate the gas exchange and pyrolytic release experiments.

The distribution and physical nature of the surface debris attracted attention from a number of workers. A map of the global distribution of different colour units identified on a planet-wide plot of red albedo versus violet albedo was derived from the Viking Orbiter approach photography (McCord *et al.*, University of Hawaii). The differences were attributed to rock-soil mixes and condensates and tentative correlations of the R-V units with units identified from telescopic spectra, which mostly resemble those of oxidized basalts, were made. A related, but more detailed, study (E.L. Strickland, Washington University) correlated colour variations seen in orbital pictures of the Chryse region with colour variations observed around Viking Lander 1, and constructed an elaborate stratigraphical sequence to explain their apparent superpositions.

Several papers were concerned with eolian processes. R.E. Arvidson of Washington University pointed out that

Michael H. Carr is in the US Geological Survey, Menlo Park, California.

despite an active eolian regime on the planet, the survival of craters on relatively ancient surfaces implies that bedrock erosion rates are extremely small ($< 10^{-7}$ cm yr $^{-1}$). The wind-induced changes observed must be due almost entirely to remobilization of relatively loose, friable debris. Wind tunnel experiments (R. Greeley *et al.*, Arizona State University) using a number of different sands and targets, which simulate wind abrasion on Mars, showed that for all conditions, abrasion rates were several orders of magnitude higher than those estimated for Mars. The results can be reconciled, Greeley suggests, if sand on Mars consists mainly of easily destroyed aggregates rather than discrete mineral grains. The aggregates probably consist of relatively soft weathered products, which disintegrate easily and have limited abrasive capability. They are nevertheless capable of saltating and forming dunes which are numerous especially at high latitudes. A somewhat different view of eolian erosion (J. McCauley, M. Grolrier and C. Breed, US Geological Survey; F. El Baz, Smithsonian Institution) is that it has taken place extensively in the high northern latitudes of Mars to produce the so-called knobby terrain. On the basis of comparisons with landforms in the deserts of south-west Egypt, they propose that the old cratered terrain was initially eroded by fluvial processes and then the interfluvies eroded away by eolian and mass wasting processes to leave the knobby remnants that we now observe.

Apart from volcanic studies and several papers on the Tharsis bulge discussed above, discussion of the bedrock geology was restricted to relatively few topics. Channels remain as controversial as when they were discovered a decade ago, and following papers by H. Masursky, M. Carr and B. Lucchitta (USGS), a lively debate took place on the origin of large channels, with water, ice and wind each receiving support. On two points a consensus is emerging; the large outflow channels (see figure on facing page) have a wide range of ages, and the branching valley networks are mostly very ancient. As to origins, it appears that most large channels are complex and a variety of processes, including catastrophic flooding, glacial scour, wind erosion and later filling by lava, may have contributed to their final form. In a related topic R. Huguenin and S. Clifford (University of Massachusetts) elaborated on their previous suggestion that water sources (oases) exist in the Solis Planum and Hellespontus regions. They suggest that continual loss of water from the regolith to the atmosphere in the equatorial regions causes a slow migration of water from the poles to the equator in the megaregolith below the permafrost. The rates are extremely slow and significant only over millions of years. The ground water is replenished by condensation of water from the atmosphere onto the poles.

Finally, several papers dealt with impact craters and what they might tell us of the nature of the martian surface. In almost all areas more than 20 per cent of fresh impact craters have central pits (W. Hale and J. Head, Brown University). These have been previously ascribed to volatiles in the surface and their ubiquity may indicate the omnipresence of ground ice. In a more elaborate study (K. Blasius and J. Cutts, Science Applications), the distribution of craters with different types of ejecta patterns was mapped and showed a complex dependence of ejecta types on geological unit, latitude and possibly other factors. Experimental simulations of martian ejecta patterns have been conducted (P. Schultz and D. Gault, Ames Research Center) and demonstrate that ejecta ramparts closely resembling those on

Mars can be reproduced by impacts into loose debris in the presence of a moderate atmospheric pressure.

Thus, in contrast to the previous Mars colloquia, emphasis of the geology papers was on surficial processes. This may simply reflect the ease with which the relevant available data can be manipulated. The infrared, colour and chemical data, upon which much of the interpretation of surficial processes rests, are readily handled in the computer. In contrast, interpretation of surface morphology, which more closely reflects the bedrock geology, largely depends on the much longer process of direct visual inspection. Analysis of the enormous volume of Viking Orbiter images to improve our knowledge of the geological evolution of the planet will clearly take many years. □

Mars climate change: where are the petroglyphs?

from Fraser P. Fanale

At one point in the colloquium it was pointed out that petroglyphs illustrating both abundant game and well preserved fragments of ostrich eggs thousands of years old could be found in almost rain-free portions of the Egyptian desert. These were considered indicative of a formerly wetter climate in that currently extremely arid region¹. By analogy, most of the papers directly aimed at the planet Mars seemed to seek some natural 'inscription' of morphology on the martian surface that might indicate a prior warmer wetter period (or, as we say, a more 'paradisical' period) in martian history. To be currently recognized, such an inscription or morphology would have to be large enough to be identifiable in available Viking Orbiter images.

The candidates most often discussed are the various channels of different sizes and detailed geometry²⁻⁴. The huge 'outflow' channels (see the figure) are spectacular but seem unconvincing as indicators of a very different past climate at the Mars surface. They lack tributaries and seem to spring at full size from what appear to be collapse features. These and other observations suggest that they burst forth from subsurface water reservoirs dammed up by permafrost⁵. Thus their relationship to global conditions in the Mars surface megaenvironment or at the atmosphere-surface interface may be subtle and indirect, or nonexistent. They could probably form today.

The so-called 'runoff' channels are valleys or gullies which are shorter, but

have well developed tributaries, form regional networks and are more suggestive of regional slow erosion by surface runoff⁶. When this geometry is examined in detail, however, it resembles that caused by headward 'sapping' processes involving subsurface water more than that expected from ordinary rainfall⁷. Nonetheless, there are two aspects of these 'runoff' channels (evidently 'network' channels would be a safer term) that persist in indicating global climate change. First, the network channels are widely believed to occur only in the oldest terrain on Mars⁶. Second, they are globally distributed within that terrain. The global distribution suggests that rainfall may have had some role even if sapping was involved as well. Moreover the restriction to the oldest (4.0 billion years?) terrain may be in harmony with theories suggesting that global conditions involving high atmospheric Pco₂ (ref. 8) or a reducing atmosphere would have to be sustained by an atypically high degassing rate, possibly associated with accretion^{9,10}, and that such conditions could not prevail for long in the face of planetary hydrogen loss and removal of CO₂ to form rock-attacking acid. It has been argued, however, that many variables (other than age) may determine channel distribution and that the restriction of any channel type to earliest Mars history has yet to be proved¹¹, while others argue a dominant role for ice^{12,13} or wind¹⁴ in carving the channels. Although a role for wind and ice in modifying the channel morphology is readily conceded, water seems to be required to explain some channel characteristics such as their reliable and familiar tendency to flow

Fraser P. Fanale is in the Planetary Geosciences Section, Institute for Geophysics, University of Hawaii.

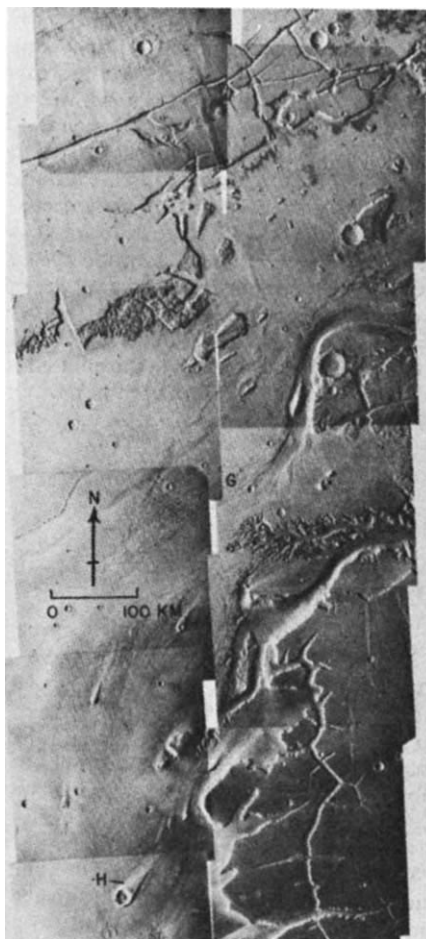
downhill¹⁵.

Besides the channels we have various morphologies indicative of the presence of subsurface H₂O ice or even liquid water. One is the so-called 'thermokarst' terrain which resembles the surface distortion and cracking patterns observed in regions of hard frozen (ice-filled) permafrost on Earth^{16,17}. Subsurface H₂O may also be indicated by 'nonlunar' or cohesive flow characteristics in crater ejecta. Most investigators concede that subsurface volatiles are involved in producing such ejecta but disagree as to how many separate categories of 'sloppy' throwout there are, whether or not some types are latitude dependent^{18,19} and which types require a specific role for water ice, liquid water, adsorbed CO₂ and atmospheric CO₂. A combined approach which studies the dependence of throwout morphology on crater size, target age and target material, together with experimental cratering in icy and muddy targets — some of which are multilayered — may allow these variables eventually to be sorted out²⁰.

Yet another 'petroglyph' is the 'layered terrain' developed near the edge of the north cap and discovered by the Mariner 1971 mission. Since its discovery, a strong connection between the Mars layered terrain and oscillatory astronomically driven climate change has been indicated^{21,22}. Such a connection, in turn, would conceptually link climate change on Mars to astronomically driven glacial cycles on Earth. Current quantitative models²³⁻²⁶ could be summarized as follows: as the obliquity (angle between rotation axis and normal to the ecliptic) of Mars increases from its present value ($\sim 24^\circ$) to its maximum (37°), the increased facing of the polar regions to the Sun will cause warming of these critical regions.

Although the large north cap is now known to be H₂O ice, not CO₂, and although the CO₂ south residual cap is almost trivial as a CO₂ reservoir, the atmospheric pressures could possibly double or even triple as a result, owing to desorption of CO₂ from large circumpolar and high latitude deposits of weathering products and pulverized igneous rock. When the obliquity falls far below its present value (to $\sim 11^\circ$) then the poles cool much more than the regolith which exists at a variety of depths and latitudes, and which experiences a far more attenuated thermal 'wave' (the equatorial regolith even warms). Hence the greatly cooling polar regions will 'cryopump' most of the CO₂ from the vast 'ocean' of adsorbed CO₂ in the regolith, resulting in the sudden appearance of a huge CO₂ polar cap (as the north cap was originally thought to be), seemingly from nowhere. In the process the atmospheric pressure falls to less than 0.5 mbar.

Thus in the models, the atmospheric pressure varies by a factor of 20 to 100 while giant CO₂ caps appear and disappear with the clock-like precision of the



Viking photomosaic of a portion of Kasei Vallis centred at approximately latitude 20°N , longitude 75°W . Outflow features such as streamline hills (H) and grooves (G) occupy the channel floor, but valley sides, tributary canyons and inner channels of the valley floor show pronounced structural control by fracture systems.

From Baker, V.R. & Kochel, R.C. *J. geophys. Res.* **84**, 7961 (1981).

obliquity cycle — a time scale of 10^5 years. Intuitively it would seem that the harmonic changes in the atmosphere's ability to transport dust and the concurrent variations in cap extent would provide ample means of producing the layered

terrain.

Current efforts to work out the mechanics of this process take advantage of the fact that the current pressure is just barely enough to raise dust. If pressure decreased by a few millibars, then dust storms would be rare; but if this pressure increased by a factor of two, they might be almost continuous²⁵.

Going back further in martian history it is thought that the obliquity cycle was of still greater amplitude before development of the huge Tharsis volcanic complex and associated gravity anomaly. This is because Tharsis near the equator actually stabilizes the planet's axial wobble²⁷; thus the effects discussed would be even more dramatic before Tharsis. Yet investigators keep insisting that the layered terrain is a feature of the latter portion of Mars' history. One solution offered at the meeting is that Tharsis was not formed at the equator, but rather flipped the planet over, causing global wandering of the poles. This could have left several prior 'layered terrains' stranded as outliers about the planet and these could, following erosion, be recognizable only as some sort of blanketing feature of which there are many²⁸. Other solutions involve a steady-state system in which layers are destroyed at the base of the piles as fast as new ones are laid down²⁹. All this provides a marvellous example of comparative planetology since the Earth's climate is affected by the same astronomical mechanism with similar periods but lower intrinsic amplitude and seemingly more complex feedback mechanisms than for Mars (see ref. 30).

The regolith-atmosphere-cap models show that even the more violent regolith-atmosphere-cap interaction before Tharsis would, by itself, be incapable of creating a sufficiently profound change in atmospheric CO₂ pressure or composition to produce a stable greenhouse system that could accommodate rainfall despite the lower energy output of the ancient Sun^{10,31}. Thus a magic additional ingredient that seemingly stabilized a more clement environment early in the history of Mars and produced the infamous channel 'networks' remains unidentified.

In summary, one is tempted to divide

1. El-Baz, F. in *J. geophys. Res.* 3rd int. Colloq. on Mars Spec. Issue (in the press).
2. Carr, M.J. *et al. Science* **193**, 766 (1976).
3. Masursky, H. *et al. J. geophys. Res.* **82**, 4016 (1977).
4. Sharp, R.P. & Malin, M.C. *Bull. geol. Soc. Am.* **86**, 593 (1975).
5. Carr, M.H. *J. geophys. Res.* **84**, 2995 (1979).
6. Carr, M.H. in *J. geophys. Res.* 3rd int. Colloq. on Mars Spec. Issue (in the press).
7. Pieri, D. *Icarus* **27**, 25 (1976).
8. Owen, T. *et al. Nature* **277**, 640 (1979).
9. Fanale, F.P. *Icarus* **15**, 279 (1971).
10. Sagan, C. & Mullen, G. *Science* **177**, 52 (1972).
11. Masursky, H. *et al. J. geophys. Res.* (in the press).
12. Lucchitta, B.K. *et al. Nature* **290**, 759 (1981).
13. Lucchitta, B.K. in *J. geophys. Res.* 3rd int. Colloq. on Mars Spec. Issue (in the press).
14. Cutts, J.A. *et al. NASA TM 79729*, 227 (1978).
15. Nummedal, D. in *J. geophys. Res.* 3rd int. Colloq. on Mars Spec. Issue (in the press).
16. Gatto, L.W. & Anderson, D.M. *Science* **188**, 255 (1975).
17. Carr, M.H. & Schaber, G.G. *J. geophys. Res.* **82**, 4039 (1977).
18. Johansen, L.A. *NASA TM 80339*, 123 (1979).
19. Cintala, M.J. & Mouginis-Mark, P.J. *Geophys. Res. Lett.* **7**, 329 (1980).
20. Horner, V.M. & Greeley, R. in *J. geophys. Res.* 3rd int. Colloq. on Mars Spec. Issue (in the press).
21. Ward, W.R. *J. geophys. Res.* **79**, 3375 (1974).
22. Cutts, J.A. *J. geophys. Res.* **4231** (1973).
23. Fanale, F.P. & Cannon, W.A. *J. geophys. Res.* **84**, 8404 (1979).
24. Toon, O.B. *et al. Icarus* **44**, 552 (1980).
25. Pollack, J.B. in *J. geophys. Res.* 3rd int. Colloq. on Mars Spec. Issue (in the press).
26. Fanale, F.P. in *J. geophys. Res.* 3rd int. Colloq. on Mars Spec. Issue (in the press).
27. Ward, W.R. *et al. J. geophys. Res.* **84**, 243 (1979).
28. Schultz, P.H. & Lutz-Garhan, A.B. in *J. geophys. Res.* 3rd int. Colloq. on Mars Spec. Issue (in the press).
29. Clifford, S.M. in *J. geophys. Res.* 3rd int. Colloq. on Mars Spec. Issue (in the press).
30. Hays, J.D. *et al. Science* **194**, 1121 (1976).
31. Pollack, J.B. *Icarus* **37**, 479 (1979).

Mars' climate history into three parts. The most recent part would be post-Tharsis (the past about 1 billion years) when obliquity variations force CO_2 exchange between the two large CO_2 reservoirs, the regolith and periodically appearing caps, through the conduit of the atmosphere. During this process the atmospheric pressure varies harmonically by more than 20-fold. The previous epoch (4.0 to 1.0×10^9 years?) saw the very same processes dominating, but with larger amplitude due to the absence of Tharsis as a stabilizing influence. Also other 'nonharmonic' or 'irreversible' trends (such as gradual changes in regolith mass, amount of

exchangeable CO_2 , solar constant) may have changed the quantitative aspects of the operation of the three-part system. Finally we have the 'networks' with their fascinating suggestions of an epoch of more clement conditions in earliest martian history. These may require feedback mechanisms such as gas or dust greenhousing or greater atmospheric heat transport to the poles, in addition to the CO_2 exchange. In any event, the networks are found (only) long ago and in another land. The issue of whether or not the wench is truly dead is one best left to the biologists who were not represented among the speakers at the colloquium. □

Paige attributes this difference to the relatively high albedo of the south cap during summer. The seasonal pattern of dust storms provides a likely explanation for cap albedo differences, dust storms being most frequent and widespread during the season of CO_2 condensation on the northern cap. Tillman's data pose a difficulty for this view, however, since they show little or no year-to-year variation in the rates of condensation or sublimation of CO_2 polar ice despite apparently large year-to-year differences in the behaviour of dust storms.

A particularly spectacular global dust storm occurred during Viking's first year. Although several explanations have been proposed for the origin of these global storms, there has been no fully satisfactory quantitative theory of their life cycle, incorporating both storm birth and death. A step in this direction may have been taken by H. Houben⁵ who used a highly simplified atmospheric circulation model to show that global-scale atmospheric disturbances can grow if the dust content is non-uniformly distributed and heating is sufficiently intense, but no global-scale disturbances can grow if the dust is uniformly distributed. Since the storms tend to produce a relatively uniform distribution of dust, this could account for storm decay without regeneration of new storms. Further study will be needed to determine whether Houben's model is applicable to the real martian atmosphere.

In another modelling study of dust storms, R. Haberle showed how dust raised in a narrow latitudinal belt, such as the southern summer subtropics, could cause a dramatic intensification and expansion of the thermally driven, zonally symmetric tropical circulation, the so-called Hadley circulation, in both hemispheres⁶. Tillman's data together with infrared data from the Viking Orbiters showed that such an expansion and intensification of the Hadley circulation actually did occur during the global dust storm observed by Viking. These observations have far-reaching implications for the relationships governing the latitudinal extent and intensity of Hadley circulations, in particular, how Hadley circulations depend on such governing parameters as planetary radius, rotation rate and heating intensity. The relationships seem to be applicable not only to the circulations of the 'sister' planets Mars and Earth, but also to the more exotic circulations at the cloud-top levels of Venus and Jupiter.

Martian atmospheric data are also helping to broaden and generalize existing theories of large-scale atmospheric waves. The travelling mid-latitude waves measured by the pressure, wind and temperature sensors at the Viking Lander sites (see the figure) have longer wavelengths

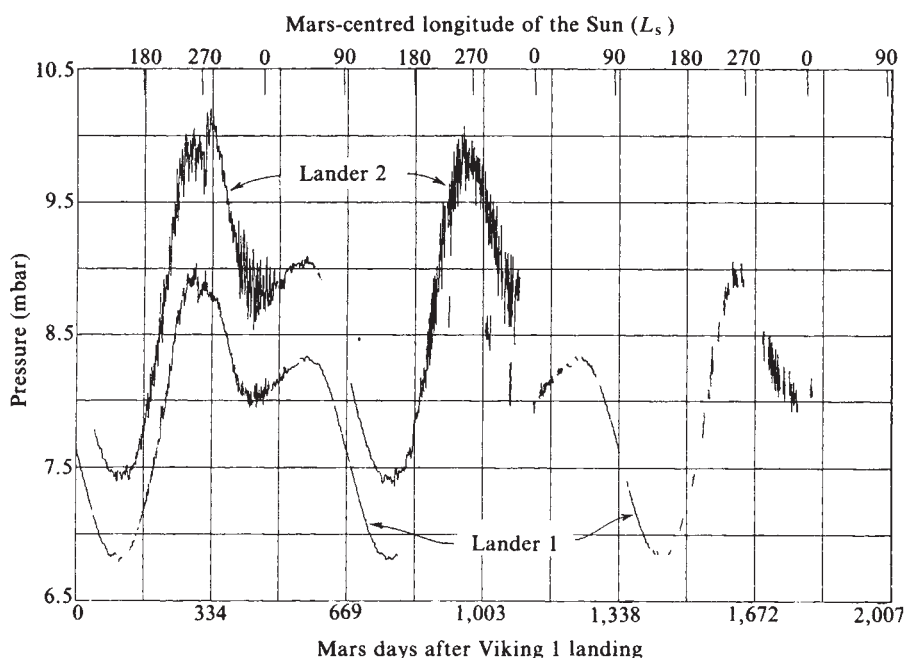
The martian lower atmosphere

from Conway B. Leovy

ALMOST three martian years (more than five Earth years) of atmospheric pressure data at two Viking Lander sites have provided a remarkably detailed picture of the weather and climate on Mars. These data, presented at the colloquium by J.E. Tillman¹, revealed the seasonal shifting of mass between the carbon dioxide polar ice caps and the atmosphere, global circulation changes associated with major dust storms, travelling waves in middle latitudes and intense atmospheric tides driven by absorption of solar radiation in the dust-laden atmosphere.

The seasonal cycling of pressure, involving as much as 30 per cent of the mass of the atmosphere, produces a massive atmos-

pheric movement back-and-forth between the two polar regions, and is indicative of the sensitivity of atmospheric mass to the polar cap radiation balance². Narumi has shown that radiation balance models can account well for the seasonal pressure variations if allowance is made for seasonally varying polar clouds³. The radiation balance is itself highly sensitive to the albedo of the polar caps, and, in a detailed analysis of Viking Orbiter data, D. Paige⁴ has shown how differences in albedos of the two caps can explain differences in their seasonal behaviour. At the end of summer, there is a small residual cap at each pole: the northern residual is H_2O ice while the southern is CO_2 ice.



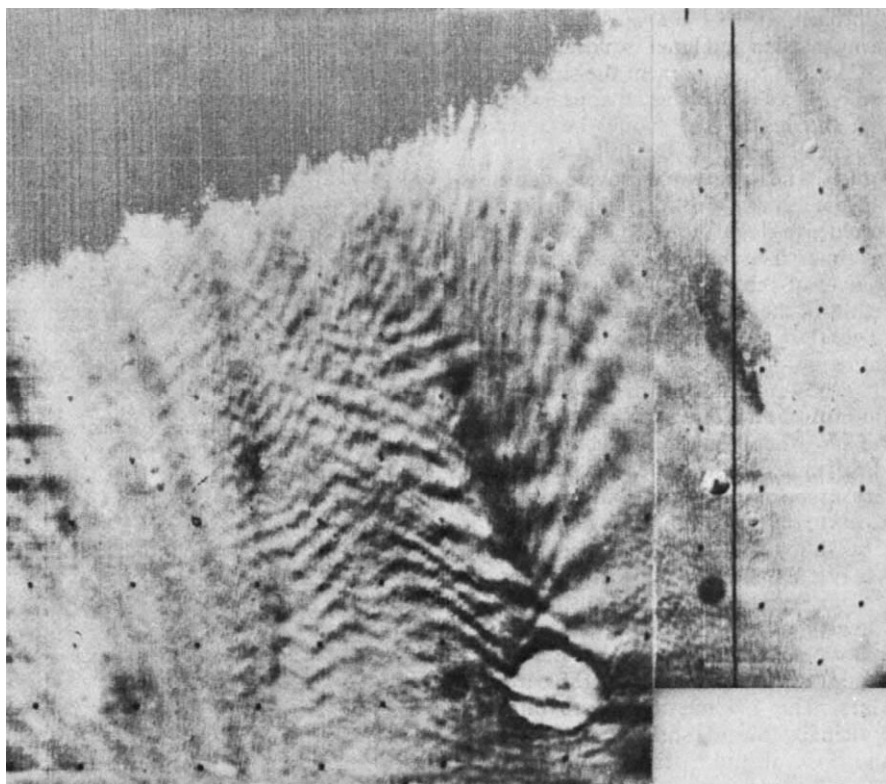
Daily averaged pressures at the Viking Landers. The ordinate interval is the number of Mars days in a Mars year. L_s values (top) correspond to season: 0, northern spring equinox; 90, northern summer solstice; 180, northern autumn equinox; 270, northern winter solstice. Gaps correspond to missing data.

Conway B. Leovy is in the Department of Atmospheric Sciences and Graduate Program of Geophysics, University of Washington, Seattle.

and are much more regular than their counterparts, the familiar travelling cyclones and anticyclones of terrestrial mid-latitudes. J. Barnes presented results of efforts to understand the wave properties theoretically⁷. The high martian static stability and the zonally aligned topography both tend to stabilize these disturbances relative to their terrestrial counterparts, and may be the causes of the great observed regularity. More generally, the scales of the observed waves and those on the other planets seem to fit in with a general scaling for planetary waves similar to that for the Hadley circulations. (See the photograph).

Aspects of the climatic history of Mars are discussed in detail in the preceding article by Fanale. The regolith and CO₂ polar ice caps seem to have had crucial roles in buffering atmospheric CO₂ pressure since the time of regolith formation. The regolith and polar caps also serve as reversible reservoirs of H₂O. Viking Orbiter measurements of atmospheric water vapour may provide an indication of the relative roles of seasonal adsorption in the regolith and condensation in the polar cap, but the interpretation of these data remains controversial. B. Jakosky and C. Farmer⁸ believe that the seasonal variations of H₂O are dominated by adsorption, whereas D. Davies and L. Wainio consider exchange with the polar cap to be more important⁹.

An important new result relevant to the earliest period of atmospheric history was a careful recalculation of the escape flux of nitrogen by J. Fox and A. Dalgarno¹⁰. Viking measurements showed an enrichment of ¹⁵N with respect to ¹⁴N by a factor of 1.7. Fox and Dalgarno were able to show that their calculated escape is consistent with this enrichment and an initial N₂ endowment of as few as ten ²²N₂ molecules per cm², much more than the current value but substantially lower than the early post-Viking estimate of McElroy *et al.*¹¹. The escape flux predicted by Fox and Dalgarno is sensitive to assumed details of the ionospheric structure which have not yet been measured, and to certain other assumptions. If the assumptions are correct, and if the N₂/CO₂ ratio of Mars is in line with that of the Earth and Venus, the Fox-Dalgarno calculation suggests a total evolved CO₂ mass equivalent to as little as 50 mbar, but it does not demand such a low value. □



Mariner 9 photograph showing a lee wave system produced by crater Milankovic, a 100-km diameter crater at 53°N, 148°W. The lee waves extend 800 km downstream to the evening terminator with a wavelength of about 60 km.

Exploration of the upper atmosphere and ionosphere of Mars

from C.T. Russell and A.F. Nagy

STRATEGIES for the exploration of the Solar System and its individual planets have been prepared by a number of different committees and boards. In recent years, the major factor restraining the pursuit of such plans has not been lack of suitable technology but of money or, more specifically, the willingness to spend that money on any particular objective. To establish a logical strategy, the sophistication of the planned investigation and the accessibility of the planet play important roles. Whether sophistication is measured in terms of the type of mission — flyby, orbiter, lander, rover or manned — or in terms of scientific objectives, the ranking of a particular mission remains roughly the same. Again, whether the accessibility of a planet is measured in terms of the required launch vehicle capability or in engineering considerations such as thermal and power requirements and mission lifetime or in available telemetry rate, one will obtain roughly the same ranking.

The cost of a mission depends on both sophistication and accessibility. For

roughly the same cost one could choose to fly a very sophisticated mission to the Moon, a modest mission to Jupiter or a very simple mission to Pluto. The total money available for the planetary program, is, however, limited and choices have to be made. It is sensible to choose a mixture of sophisticated and simple missions, because detailed data obtained from the more accessible bodies can help us to understand the behaviour of the more inaccessible objects, and comparisons between less sophisticated and more fragmented data obtained about a group of planets can be used to learn more about the generalities of planetary evolution. This approach is usually adopted because budgets are tight, and mission planners must seek missions that provide new and exciting data at the minimum cost. However, there are exceptions to this

C.T. Russell is in the Institute of Geophysics and Planetary Physics, University of California, Los Angeles, and A.F. Nagy is in the Department of Atmospheric and Oceanic Science, University of Michigan, Ann Arbor.

1. Tillman, J.E., Henry, R.M. & Hess, S. *J. geophys. Res.* **82**, 2947 (1979).
2. Leighton, R.B. & Murray, B. *Science* **153**, 136 (1966).
3. Narumi, Y. *Proc. 12th Lunar and Planetary Symp.*, 234 (1979).
4. Paige, D.A. *Abstr. 3rd int. Colloq. on Mars. Jet Propulsion Laboratory, Pasadena*, p.179 (1981).
5. Houben, H. *ibid.*, p.117 (1981).
6. Haberle, R.M. thesis, Univ. Washington (1981).
7. Barnes, J.R. *J. Atmos. Sci.* **38**, 225 (1981).
8. Jakosky, B.M. & Farmer, C.B. *J. geophys. Res.* **86** (in the press).
9. Davies, D.W. & Wainio, L.A. *Icarus* **39** (in the press).
10. Fox, J.L. & Dalgarno, A. *Planet. Space Sci.* **28**, 41 (1980).
11. McElroy, M.B., Kong, T.Y., Yung, Y.L. & Neir, A. *Science* **194**, 1295 (1976).

approach, two of the most major ones being martian and lunar exploration.

Consider for a moment the strategy for exploring a single planet. It appears logical that the initial step should be a reconnaissance with a flyby followed by an orbiter. The flyby would provide snapshots of some characteristics, and the orbiter would give global data over a longer period of time. The instrumentation and objectives of the orbiter will depend on the results of the flyby. It is not obvious that a single flyby or a single orbiter can do the entire job — orbits suited for one objective may be unsuitable for another, for example. Orbiters certainly provide increased observation time, but may also make more sophisticated and/or global measurements possible. Landers provide access to the surface and a longer period for observation and analysis, although at only one site. Sample return provides access to yet more sophisticated analysis as well as allowing longer time for analysis.

Both the United States and the Soviet Union have orbited spacecraft around Mars. The US missions were directed principally towards surface, photogeology and low altitude atmospheric science objectives. The Soviet missions looked at the solar wind interaction with Mars but telemetered little data and had relatively high minimum altitudes that did not penetrate the upper atmosphere and ionosphere of Mars. Hence there is only very limited information on the nature of the upper atmosphere and ionosphere of Mars, little geophysical or geochemical data, only a rough idea of the topography and an essentially unresolved controversy on the existence of an intrinsic planetary magnetic field. Furthermore, the proper global geochemical and geophysical data to plan an intelligent surface exploration strategy are not currently available. The US skipped a stage of exploration in order to be first to return data from the surface. The US and European scientific communities have become aware of this void and are coming forward with proposals to fill the gap. An ESA mission, KEPLER, has been proposed but unfortunately it competes with the Polar Orbiting Lunar Observatory (POLO), also in the planning process at ESA, which is designed to fill the analogous gap in our lunar knowledge. In the US, a Pioneer class Mars Orbiter mission, in many ways similar to KEPLER, has been discussed, but the chances for such a mission are very uncertain given the present tight financial situation at NASA.

In view of this lack of data there was little discussion of martian aeronomy or the solar wind interaction at the conference. J.A. Slavin (UCLA) addressed the existence of a planetary magnetic field. The few cases in which the Soviet Mars orbiters were claimed to enter either a martian magnetosphere or magnetotail are subject to varying interpretations. How far does the bow shock stand off of the planet and what size object would be necessary to create a



Photograph of Mars taken from a distance of 419,000 km by Viking 2. Near the top is the volcano Ascraeus Mons, with a white water-ice cloud to the north-west. The Valles Marineris is visible below the volcano, and the impact crater of the Argyre Basin, covered with winter frost, at the bottom of the photograph. The picture below shows the horizon of Mars with clouds, thought to be crystals of CO_2 , 25–40 km above the surface. Taken 11 July 1976 by Viking Orbiter 1.



bow shock in this position? Slavin showed that a size slightly larger than the planet plus its ionosphere is necessary and thus he feels there is a weak intrinsic field with a moment of about $1.4 \times 10^{22} \text{ G cm}^{-3}$, about 5,000 times smaller than the Earth's moment. Examining available ionospheric data, principally from radio occultations, he finds no evidence for a Venus-type interaction. On the other hand, at altitudes above the ionosphere there are similarities in that heavy ions are found to be carried

away by solar wind from both planets.

The lack of martian aeronomical data is particularly intriguing to planetary scientists because of the contrasts between Venus, Earth and Mars. Earth and Mars have similar rotation periods. Venus and the Earth have similar gravitational fields. Mars and Venus have similar atmospheric composition. Venus has no — or at least no detectable — planetary magnetic field while Mars has at most a weak one. In view of these contrasts, will the martian thermospheric temperature be controlled by solar EUV like the Earth's or be rather insensitive to solar EUV like that of Venus? Does Mars have a strong upper atmospheric wind system like Venus?

The ionosphere of Mars is similarly an enigma. Is there a solar wind source of heat as on Venus? What maintains the nightside ionosphere revealed by radio occultation? The terrestrial ionosphere is partially maintained at night by the large reservoir in the magnetosphere that fills with plasma during the day and empties at night. On Venus the night ionosphere is maintained by rapid flows from the dayside and by energetic particle bombardment from the wake of Venus. Finally, the chemistry of the martian ionosphere is a virtual unknown. The neutral atmospheric composition and hence the ionospheric composition are expected to be similar to that of Venus. However, the reaction rates are temperature and velocity dependent and therefore quite a few surprises may be in store.

Venus was found to have an extended non-thermal hydrogen and oxygen corona even though it has a very low exospheric temperature; the presence of this corona is due to dissociative recombination and charge-exchange reactions. On the other hand the Earth has an extensive hydrogen corona caused by the Jean's escape flux and charge-exchange processes, but it does not have a significant oxygen corona. The solar wind is deflected by the terrestrial magnetosphere so there is little significant interaction with the geocorona, but in the case of Venus, with no intrinsic field, the situation is very different, in that there is significant absorption by the corona through ionization in and charge-exchange with the flowing solar wind. In this regard Venus acts very much like a comet. Does Mars behave the same way, or is its corona shielded from the solar wind by a small magnetosphere?

These questions and many more must wait until we get back to a logical series of explorations, for despite the sophistication of some of our knowledge about Mars we are completely ignorant about some of its basic features. Our best hope is ESA's KEPLER or NASA's Pioneer Mars Orbiter missions. The realization of either of these missions would provide a big step forwards in our knowledge of Mars, as well as exposing many areas of ignorance to stimulate further progression in our exploration of the planets. □

REVIEW ARTICLE

The role of gene dosage and genetic transpositions in carcinogenesis

George Klein

Department of Tumor Biology, Karolinska Institutet, S-104 01 Stockholm, Sweden

Analysis of chromosome translocations associated with B-cell derived tumours and studies of chromosome-15 trisomy in murine leukaemia support the theory, emerging from recent work on tumour viruses, that cancer can be associated with DNA rearrangements which result in the increased expression of normal cellular genes.

IN a recent review on the origin of human cancers, Cairns¹ suggested that most human cancers result from genetic transpositions, rather than mutations. The purpose of the present discussion is to review some of the evidence dealing with the role of specific translocations in the genesis of certain human and animal tumours. Recent findings on mouse plasmacytoma and human Burkitt's lymphoma suggest that the translocations frequently associated with these tumours may act through the activation of specific gene(s) that regulate the growth and differentiation of the lymphocyte-plasmacyte series. It is suggested that translocation of the relevant chromosome segment to an immunoglobulin determinant-carrying region brings the gene(s) under the influence of a highly active cellular promoter, leading to increased expression of the gene product. It will be further argued that increased gene expression is the cause of murine leukaemia associated with trisomy, although in that case the gene may also be altered by mutation or viral promoter insertion. The general theme, that cell transformation is the result of an increase in the expression (or functional activity) of a normal cellular product, will be considered in the light of existing knowledge about the action of virally transmitted, cell-derived oncogenes.

Viral oncogenes

Retroviruses: The rapid expansion of viral oncology during the 1960s was considerably motivated by the expectation that specific, virally coded genes determine the synthesis of transforming proteins. This turned out to be true in one sense, but the situation is quite different from what was expected. In several cases, the transforming proteins associated with the directly acting retroviruses are similar to normal cellular constituents, amplified in amount and/or changed in functional activity. All viruses of this category share the following properties:

- First, they can induce tumours in susceptible hosts *in vivo* after short latency periods.
- Second, they transform cells *in vitro*.
- Third, the viral genome carries sequences that appear to have been derived from the normal host genome ('oncogenes'). In most cases, they are inserts that replace large portions of the viral genome. Consequently, the virus becomes defective and requires a helper for replication. Retrovirus-carried transforming sequences are known to be present in all normal cells of the species of origin; corresponding, highly conserved sequences were found in foreign and sometimes quite distant species (see ref. 2). Different transforming retroviruses of the same species often carry partially or wholly different cellular sequences³, suggesting that specific cell-derived sequences could be related to specific forms of transformation³.
- Fourth, retrovirus-carried oncogenes code for transforming proteins, now well defined in several systems. The pp60^{src} of Rous sarcoma virus (RSV)⁴ and the p120 of the Abelson virus⁵ are among the best known examples. Although they are very

different proteins, they share an unusual protein kinase activity, leading to the preferential phosphorylation of tyrosine. It has been suggested that they may act by modifying one or more cellular polypeptides⁶, the identity of which is being extensively investigated. At least a portion of the pp60^{src} and the Abelson p120 protein is associated with the cell membrane; proteins of the cytoskeleton are among the obvious candidates (for review see ref. 7).

Although intriguing, tyrosine phosphorylation cannot be regarded as a universal mechanism of transformation, as mouse cells transformed by Kirsten and Moloney sarcoma virus or SV40 virus show no increased abundance of phosphotyrosine⁶. Also, there is no evidence that the transforming proteins of the defective avian leukaemia viruses (DLVs) have any kinase activity. These agents have been divided into three classes⁸: erythroblastosis, myelocytomatosis (*in vitro* macrophage transforming) and myeloblastosis-type viruses. Some of these viruses have other transforming effects in addition to that indicated by their name (for details see ref. 3). The oncogenes of the three virus groups, designated *erb*, *mac* and *myb*, represent sequences that have been highly conserved through the phylogeny of higher vertebrates and they are unrelated to the *src* gene of RSV³. They are transcribed in normal cells, but at a much lower level than in transformed cells. There is a strict correlation between the presence of *erb*, *mac* and *myb* sequences, and the ability of the virus to transform erythroblasts, macrophage-like cells or myeloblasts. Studies using temperature-sensitive mutants⁹ have shown that a virus-encoded protein is required to maintain the undifferentiated state of transformed erythroblasts. It has been suggested that the DLV-carried oncogenes represent normal cellular genes, coding for lineage-specific proteins involved in the control of haematopoietic differentiation. This would explain the target cell specificity of the DLVs, and suggests that the oncogene product has to be expressed continuously to maintain the differentiation block.

Recently, Hayward *et al.*¹⁰ found that the slow-acting, non-defective, non-transforming avian leukosis virus (ALV) activates the cellular counterpart of the virally transmitted *myc* sequence, *c-myc*. Like other proviruses, ALV integrates randomly with the host cell genome, but occasionally it becomes inserted immediately upstream from the *c-myc* gene. The viral promoter, included in the long terminal repeat (LTR) sequence at the 3' end of the provirus, has been shown to be responsible for the enhanced expression of *c-myc*. RNA transcripts have been identified that contained both viral LTR- and *c-myc*-encoded information and are present at levels 30- to 100-fold higher than the corresponding *c-myc* RNA of the normal cell.

The 'promoter insertion model', proposed on the basis of these findings, is potentially of great interest. It suggests that a non-transforming, chronically acting virus may cause leukaemia by altering the expression of a normal cellular gene. It also raises the question of whether the activation of *c-onc* genes may

provide a common denominator between the carcinogenic action of viral and non-viral agents. I shall return to this topic in a later section dealing with the tumour-associated translocations.

The examples considered so far all suggest that overproduction of a normal cellular protein, rather than the appearance of new proteins that specifically cause transformation, is the mechanism of retroviral transformation. Is it now possible to disregard the main alternative explanation that postulates the existence of small differences between the viral oncogene product and the corresponding normal cellular product? This can only be answered from a few cases at present. The most convincing evidence comes from the joining of the normal cellular counterpart of a murine sarcoma virus oncogene (*mos*) with the flanking viral sequences required for infection, transcription and integration. Transfection of appropriate target cells resulted in typical transformed foci^{11,12}, supporting the view that retroviral transformation is not due to qualitative differences between the cellular and viral oncogene products. This conclusion was further supported by the finding that the transforming gene of a mouse sarcoma virus shows almost complete sequence homology with its cellular homologue¹³.

If it is correct that the viral oncogene products are closely similar or identical to corresponding normal cellular proteins, but made in excessive quantities and/or in the wrong place or time, a variety of mechanisms may be explored to explain their actions. All the mechanisms start from the thesis that host-cell feedback regulations are delicately balanced and may be disrupted by relatively small changes. This may be only the first step towards fully autonomous tumours. Most naturally occurring tumours evolve by stepwise progression, and virally initiated tumours are no exception¹⁴. Autonomy, the degree of independence from growth controlling factors, is not an absolute notion—the initial transformation event may represent only a first step, and sequential modifications may lead to increased tumorigenicity.

DNA tumour viruses: These differ from the retroviruses in some fundamental respects. Whereas the retrovirally transmitted oncogenes represent an unnecessary burden as far as the viral cycle is concerned, the transforming information of the DNA tumour viruses is essential to the viral replication cycle. Moreover, while the productive cycle of the RNA tumour viruses does not interfere with cell multiplication, DNA virus replication is incompatible with cell survival, and as a consequence, the transforming information must be an early product of the viral genome. Considerable progress has been made in identifying the transforming proteins of the small papovaviruses, SV40 and polyoma.

In the SV40 system, large T antigen is the main candidate transforming protein, alone or with small T ('small t') antigen^{15,16}. The understanding of its function is complicated by the fact that it is also involved in regulating the late viral replication cycle. Large T binds to the origin of viral DNA replication and also inhibits the synthesis of early mRNA. In microinjection experiments, a critical amount of large T antigen was required before viral DNA replication and the synthesis of virion antigens could take place¹⁷.

How can a product that must have evolved to promote the late part of the viral cycle, also be a transforming protein? One clue may be provided by the important observation of Lane and Crawford¹⁸ that SV40 large T antigen is complexed to a cellular 53,000 molecular weight (53 K) protein in the nucleus of the transformed cell. A closely similar, and probably identical, 53 K protein has subsequently been found in a wide variety of transformed and tumour cells, including chemically induced mouse sarcomas, spontaneous teratocarcinomas, Abelson virus induced B-cell leukaemias, adenovirus-transformed cells and a variety (but not all) of other transformed or tumour-derived cell lines of murine and human origin^{19–27}. Small amounts of 53 K protein were also detected in some normal mouse tissues²⁰. A comparison of turnover rates suggests that the role of the SV40 large T antigen in transformation may relate to its ability to

stabilize 53 K²⁰, leading to constitutively high expression. The ability of large T antigen to promote the late viral cycle would be rendered unimportant by the fact that transformation usually occurs in non-permissive cells. The ability of large T to trigger host cell DNA synthesis may also be important¹⁵.

The proposed role for the large T antigen–53 K complex in transformation received unexpected independent support from studies of an entirely unrelated transforming DNA virus system, the Epstein–Barr virus (EBV). The only virally induced product that is regularly found in EBV-transformed cells, the nuclear antigen (EBNA)²⁸, was found to consist of a virus-specific 48 K subunit, complexed with 53 K of cellular origin^{29,30}. A similar (probably identical) 53 K protein, not complexed to 48 K, was also present in some EBV-negative B-cell-derived lymphoma lines, but in smaller quantities than in the EBV-carrying lines. EBV conversion of these lines into stably EBNA-positive variants increased their content of 53 K protein³⁰.

The way in which EBV results in transformation is virtually unexplored. As EBNA is always associated with transformation and is the only virally determined product known to be expressed in transformed cells, it is likely to have a key role. Recently, non-permissive cells devoid of EBV receptors have been successfully infected with EBV by microinjection³¹ or by implantation of EBV-receptor-carrying membranes^{32–34}. In such cells, little or no EBNA is produced and the virus enters directly into the lytic cycle. It is thus intriguing to speculate that the appearance of the EBNA–53 K complex is involved in transformation and may even prevent the cells from entering the lytic cycle. To further explore the possible role of 53 K in transformation and/or tumour progression, comparison of cell populations that have reached different stages in progression will be of considerable interest.

One such system can be provided by infecting certain EBV-negative human B-cell-derived lymphomas with EBV. This leads to a decreased serum requirement, increased resistance to saturation conditions, changed lectin agglutinability and capping behaviour and increased cloning efficiency in soft agarose^{35–40}, and is accompanied by a parallel increase in the amount of 53 K³⁰. As another example, the serial propagation of SV40-transformed cell lines, an exercise in *in vitro* progression, tends to favour the selective amplification of the DNA sequences coding for large T antigen^{15,41}. As the amount of large T antigen and 53 K protein were found to vary in parallel^{15,18,20,42}, the corresponding increase in the amount of the 53 K protein may have an important selective role.

Despite their extensive functional and structural homology, the transforming functions of SV40 and polyoma virus seem to be very different. For polyoma, there is extensive evidence from both genetic and DNA transfection studies that large T is neither necessary nor sufficient for transformation⁴³. In all probability, middle T, alone or together with small T, is responsible for the transforming function. Middle T is a 56 K protein coded by the early region of the viral genome. It shares its N-terminal region with large T but is otherwise very different, as a result of splicing of the mRNA precursor and subsequent frameshift. Like pp60^{src}, middle T binds to the inner surface of the cell membrane and is associated with protein kinase activity^{43,44}. There is no corresponding antigen in the SV40 system.

It has been suggested that part of the middle T sequence has originated from the host genome⁴⁵, but the evidence is inconclusive. Polyoma middle T therefore still remains the most likely candidate for the role of a virally coded transforming protein. This conclusion is also supported by the fact that while polyoma-transformed cells contain the cellular 53 K protein¹⁸, it is present in an ~10 times lower concentration and is not bound to polyoma T antigen (L. Crawford, personal communication).

In conclusion, the two best known transforming DNA tumour viruses, polyoma and SV40, use very different strategies in transforming cells. For polyoma, there is a possible analogy with the transforming retroviruses insofar as transforming middle T protein has kinase activity and is associated with the inner part of the plasma membrane. For SV40, a very different mechanism,

stabilization of the transformation-related cellular 53 K protein by large T, may have a crucial role, particularly since a very similar complexing and stabilization occurs in an unrelated, but highly transforming herpesvirus, EBV.

Trisomy in murine leukaemia

All known murine leukaemia viruses, except the Abelson virus, are non-defective, lack direct transforming activity *in vitro*, carry no cell-derived sequences and induce leukaemia only after long latency periods. Their oncogenic effect is often dependent on prolonged viraemia and is heavily influenced by the genetics of the host⁴⁶.

Most virally and non-virally induced murine T-cell leukaemias are chromosomally abnormal. Trisomy of chromosome 15 is the dominating and often the only change⁴⁷⁻⁵⁰. Other trisomies (for example, trisomy 17) occur more rarely and usually represent an additional change to trisomy 15^{47,50}. Trisomy 15 has also been found in B-cell lymphomas, often together with other chromosomal changes^{51,52}. Chromosome 15-trisomic leukaemias have been induced by Moloney virus^{49,53}, radiation leukaemia virus⁴⁷ (Fig. 1), X-ray radiation⁵⁴ and various chemical carcinogens^{48,55,56}. Most spontaneously arising AKR leukaemias were also trisomic for chromosome 15⁵⁷. The cytogenetic analysis of T-cell leukaemias induced in mice carrying various types of chromosome 15 translocations has led to the conclusion that the critically important region that tends to be duplicated in the course of leukaemogenesis is localized in the distal region of the chromosome^{49,58}.

Is the duplication of chromosome 15 a secondary consequence of leukaemogenesis, or is it directly involved in causing the transition from the preleukaemic to the leukaemic state? According to the former alternative, it could be imagined that only 15-trisomic leukaemia cells can survive among all trisomies generated by non-disjunction whereas the others are lethal. This possibility could be excluded by banding analysis of chemically (dimethylbenz(a)anthrene) and virus (Moloney) induced T-cell leukaemias, derived from three different mouse stocks, carrying different Robertsonian translocations⁵⁰. The translocations were derived from the centromeric fusion of chromosome 15 with another large autosome (chromosomes 1, 5 and 6, respectively). The trisomic leukaemias contained three copies of the translocation chromosome, in most cases the only detectable chromosomal anomaly. This can only mean that the leukaemia-associated duplication of chromosome 15 is an important event in leukaemogenesis, as it compels the centromerically fused large autosome to become a 'fellow traveller'. It also shows that the trisomy of the attached autosome is by no means lethal for the leukaemia cell.

How does trisomy 15 contribute to leukaemogenesis? Consider the AKR mouse strain, inbred and selected for high leukaemia incidence. At least four independent genetic systems are known to favour leukaemia development in this strain: (1) the integrated ecotropic (Gross-subtype) proviral DNA, localized at two sites in the genome, Akv-1 and Akv-2 (ref. 46); (2) the Fv-1^a amplification system that promotes virus production⁵⁹; (3) the H-2-linked Rgv-1^s gene⁶⁰, believed to be responsible for the deficient immune response of the AKR mouse to the leukaemia-associated antigens; and (4) a less clearly defined system that increases the likelihood of the leukaemic change directly at the level of the T cell^{61,62}. In spite of these multiple leukaemia-favouring mechanisms, leukaemias appear only after a long latency period, in middle-aged or old mice, and are frequently trisomic for chromosome 15⁵⁷.

One reason for the long latency period may lie in the requirement to generate recombinants between eco- and xenotropic C-type virus. Such recombinants are believed to bear a more direct responsibility for leukaemia induction than the ecotropic virus by itself⁶³. The generation of trisomy-15 by random non-disjunction may be another time-consuming process. The promoter insertion model of Hayward *et al.*¹⁰ may be considered as a third possibility. These alternatives are not



Fig. 1 G-banded karyotype and metaphase plate of leukaemia cell originating from the thymus of a male C57BL/6 mouse inoculated intrathymically with A-RadLV (radiation leukaemia virus)-induced preleukaemia cells. Note the presence of trisomy 15 (arrows).

mutually exclusive; they may represent complementary events, acting synergistically or in succession. One may envisage a broadening of the host range by recombination, leading to the reinfection of a large number of potential target cells. Random integration of the reinfecting viral genome would have no leukaemogenic consequence until the appropriate oncogene happens to be activated by the rare event of integration of the viral promoter in its neighbourhood. Duplication of the chromosome that carries the activated oncogene would help to overcome the influence of a *trans*-acting regulator, coded by the normal homologue, and experimental evidence supporting this is summarized below.

Genetic studies on the mechanisms of murine leukaemogenesis have shown that part of the genetic variation between high and low leukaemia strains acted at the level of the target T lymphocyte itself. This information came from experiments where normal thymus tissue from a high and a low leukaemia-prone strain was transplanted to the common, thymectomized F₁ hybrid host. Despite the fact that they were in a common host environment, exposed to the same immunological and viral influences, the thymocytes of the high leukaemia strain became leukaemic more frequently than did the cells of the low leukaemia strain⁶¹.

Analysis of trisomy 15 in leukaemias arising in F₁ hybrid hosts, and related somatic hybrid studies, suggest that chromosome 15-localized gene(s) may influence the likelihood

of transformation at the target cell level. We have found^{55,56} that T-cell leukaemias induced by dimethylbenz(a)anthracene or Moloney virus in F₁ hybrids derived from crosses between two mouse strains with cytogenetically distinguishable 15-chromosomes, show a highly asymmetrical trisomy. In most crosses, one of the two parental 15-chromosomes was duplicated preferentially, depending on the strain combination.

For a given genotype, the preference was the same for the virally and the chemically induced leukaemias. In crosses where the difference between the parental 15-chromosomes was only morphological but not genetic (for example, CBA × CBAT6T6 or AKR × AKR Rb6; 15) there was no significant preference, showing that the genetic content, rather than the translocated state of the chromosome, was responsible for the asymmetry. Chromosomes 15 derived from different mouse strains showed a distinct hierarchy with regard to leukaemia-associated preferential duplication. As non-disjunction of the two homologous 15-chromosomes must occur with equal frequency, this can only be interpreted to mean that the 15-chromosome-associated gene(s) that influence leukaemia development is subject to a certain variation between different mouse strains. Duplication of one of the chromosomes (in a heterozygote) is more likely to cause leukaemia than duplication of the other.

Somatic hybrid studies⁶⁴ have provided further evidence for the existence and probable nature of this variation. A 15-trisomic AKR-derived lymphoma, designated TIKAUT, was fused with normal CBA T6T6 fibroblasts, carrying a 14;15 translocation. Resulting highly tumorigenic segregants contained the tumour-derived chromosome 15 in five to six copies, as a rule, instead of the expected three copies. The normal parent-derived chromosome 15 was only present in one copy. Segregants of low tumorigenicity usually contained only two tumour-derived 15-chromosomes and two normal-derived 15-chromosomes. Other autosomes showed only minor random variations around the expected number, four. This suggests that the normal and the tumour-derived 15-chromosomes differ in their genetic content; the former has elements that favour and the latter that inhibit tumorigenicity in the syngeneic host. We must postulate that the duplicated chromosome 15 in the original tumour, as well as the amplified chromosome 15 in the somatic hybrid, carries a mutated form of a gene that normally controls lymphocyte growth and differentiation. Alternatively, the same gene may have been permanently activated by proviral DNA insertion. The phenotypic effects of this change would not be expressed, however, unless the chromosome is duplicated by non-disjunction, needed to overcome the suppressive influence

of a *trans*-acting regulator, produced by the normal homologous chromosome. This hypothesis can be tested by cytogenetic analysis of segregants of hybrids between 15-heterozygous trisomic leukaemias and normal cells. If the hypothesis is correct, only the duplicated chromosome will be amplified in highly tumorigenic segregants.

Translocations

While the vast majority of human and animal tumours fail to show specific translocations, highly regular translocations are associated with chronic granulocytic leukaemia in humans (Ph₁ chromosome), with Burkitt's lymphoma and with murine plasmacytomas. This review has been limited to the latter two, B-lymphocyte-derived tumours.

Two characteristic translocations have been found in mineral-oil-induced murine plasmacytomas^{65,66}. The more common marker, seen in both κ and λ producers, arises by the translocation of the distal part of chromosome 15 to chromosome 12, known to carry the heavy chain immunoglobulin locus (Fig. 2). The breakpoints are D3 in chromosome 15 and F2 in chromosome 12. The heavy chain immunoglobulin loci are located proximally to band F2⁶⁷.

A reciprocal 6;15 translocation was also seen but only in κ chain-producing plasmacytomas^{65,66} (Fig. 3). Chromosome 6 is known to carry the κ chain locus⁶⁸, although its precise location is unknown. The breakpoint on chromosome 15 was in band D3, as in the 12;15 translocation.

In diploid or near-diploid tumours, the translocations only affected one of the two homologous chromosomes. In tetraploid plasmacytoma cells, the translocation must have occurred before polyploidization, because two of the four homologous chromosomes showed the translocation.

In each plasmacytoma clone only one of the two homologous immunoglobulin loci is responsible for the synthesis of a complete immunoglobulin molecule. It is of great interest to explore whether or not the translocation selectively affects the active or the inactive chromosome 12. This can be done by correlated cytogenetic and immunoglobulin allotype studies on plasmacytomas heterozygous for the allotypic marker and for an appropriate Robertsonian translocation.

The segment of chromosome 15 involved in the plasmacytoma-associated translocations is localized within the region that becomes regularly duplicated in the 15-trisomic B- and T-cell leukaemias (including cases where parts of chromosome



Fig. 2 G-banded karyotype of a hypertetraploid plasmacytoma cell. Note the presence of two copies of the T(12;15) translocation and two deleted 15 chromosomes (del 15).



Fig. 3 G-banded karyotype of a hypertetraploid plasmacytoma cell. Note the presence of two copies of the reciprocal translocation between chromosomes 6 and 15 (rcpT (6;15)).

15 are translocated to other chromosomes)^{49,58}. It is likely that this region carries gene(s) that control lymphocyte-plasmacyte differentiation and growth. Translocation to the immunoglobulin gene region, an area that is highly active in B lymphocytes and plasma cells, may bring the gene(s) under the influence of a highly active promoter. This could lead to the overproduction of some important regulator, in a manner analogous to retrovirally mediated oncogene transmission and the viral promoter model. It is important to explore whether the breakpoint on the immunoglobulin-locus-carrying chromosome affects the same nucleotide sequences in different tumours and how it is related to the immunoglobulin locus itself.

If studies of this type confirm that the translocation affects a specific site within or near the immunoglobulin locus, one may still ask whether this is due to some special vulnerability of this region to breakage, due to the DNA rearrangements that are known to occur in the area. This is unlikely, however, because λ -producer plasmacytomas showed only 12;15 but no 6;15 translocations⁶⁹, even though the κ region is known to be deleted or rearranged in λ producers⁷⁰.

The promoter activation hypothesis also implies that the translocation of the distal part of chromosome 15 to the immunoglobulin-locus-carrying chromosomes, rather than the deletion from 15, is responsible for the malignant transformation. This hypothesis can be tested by parallel chromosome segregation-tumorigenicity studies on appropriately constructed somatic hybrids.

A possibly analogous translocation has been discovered in a human B-cell-derived tumour. Both the EBV-negative and the EBV-carrying forms of Burkitt's lymphoma were found to carry a highly specific reciprocal 8;14 translocation⁷¹⁻⁷³. Banding analysis showed the involvement of the same breakpoints in different individual cases, affecting bands 8q24 and 14q32. The same translocation was also found in B-cell-derived acute lymphocytic leukaemia^{74,75}, believed to originate from the same type of target cell as Burkitt's lymphoma. Other lymphomas were also found to carry 14q+ markers in a considerable number of cases, but the variability of the breakpoints and of the donor chromosome distinguish these markers from the specifically Burkitt's lymphoma-associated 8;14 translocation⁷⁶.

Recently, variant forms of the Burkitt's lymphoma translocation have been described. In these cases, the same distal piece of chromosome 8 becomes translocated, but to chromosome 2 or 22 rather than chromosome 14⁷⁷⁻⁸¹. The situation is reminiscent of the Philadelphia chromosome, where

a 22;9 translocation is found in most cases, while a minority (~10%) shows a 22-deletion with the same breakpoint, but with either no identifiable translocation site or a translocation to other recipient chromosomes⁸².

The variant Burkitt's lymphoma translocations suggest that the distal part of chromosome 8, rather than chromosome 14, carries the gene(s) responsible for the development of Burkitt's lymphoma. As chromosome 14 is known to contain the immunoglobulin heavy chain loci in man⁸³, the human 8;14 translocation may be functionally analogous to the murine plasmacytoma-associated 12;15. This concept is strongly supported by the recent demonstration⁸⁴ that the human κ gene is on chromosome 2, and the λ gene on chromosome 22—the chromosomes involved as recipients of the distal chromosome 8 fragment in the variant Burkitt's lymphoma translocations. Interestingly, the κ gene has been localized to the short arm of chromosome 2, in the same region as the breakpoint involved in the 2;8 translocation.

Perspectives

Two major alternatives are now under consideration to explain the action of the virally transduced, cell-derived oncogenes. According to one hypothesis, transformation is merely a matter of gene product dosage. Due to the integration of the oncogene in the wrong place or in the wrong cell, a certain normal cellular product is synthesized in excessive quantity or at the wrong time, interfering with the normal differentiation programme. The second alternative predicts the existence of subtle qualitative differences between the virally carried oncogene and the corresponding normal cellular gene.

While the problem cannot be regarded as definitely settled, the recent successful transformation of fibroblasts with the normal cell-derived counterpart of the murine sarcoma virus-associated oncogene^{11,12}, and even with randomly sheared DNA from normal cells⁸⁵⁻⁸⁷ argues against the second hypothesis.

For tumour-associated non-random chromosomal changes such as chromosome 15 trisomy it is simplest to invoke gene dose effects. It may be objected that a change in gene dosage of 50% is a relatively minor change. On the other hand, the relationship between the normal cell and its many growth controlling signals must be extremely finely poised, so that small changes may tip the balance and favour unlimited proliferation. However, the chromosomal analysis of high and low tumorigenic segregants of somatic cell hybrids indicates that there also

exists a qualitative difference between the relevant gene(s) carried by the tumour- and the normal cell-derived chromosome 15. This is probably due to a mutation or some other change, like proviral DNA insertion. Duplication of the altered chromosome has probably overcome *trans*-acting regulation by the normal homologue, resulting in the leukaemic phenotype.

Both quantitative and qualitative effects must be considered in relation to the murine plasmacytoma-associated 12; 15 and 6; 15 translocations. A simple transposition of the distal region of chromosome 15, presumed to be involved in the control of lymphocyte-plasmacyte growth and differentiation, to the highly active immunoglobulin-producing gene region would be in line with the gene product dosage hypothesis, provided it can be shown that the translocation involves the active and not the allelically excluded immunoglobulin-producing chromosome and takes place within or close to the immunoglobulin gene region. Alternatively, it is possible that sequence changes ('mutations') are also involved here. Somatic hybridization-segregation experiments involving parallel cytogenetic and tumorigenicity tests of the same type as on the trisomic lymphomas may give decisive information on this question. The Burkitt lymphoma-associated specific translocations, where the distal fragment of chromosome 8 is translocated to the heavy chain immunoglobulin-cluster-carrying chromosome 14 in typical cases or to the κ light chain gene-carrying chromosome 2 or the λ light chain gene-carrying chromosome 22 in variants, lends strong independent support to the hypothesis. Yet further support comes from the fact that the κ gene is localized in the short arm of chromosome 2, which is precisely where the translocation occurs (band p12).

The constitutive switch-on of cellular oncogenes is emerging as a fairly common theme in carcinogenesis. The evidence is far

from complete and there are already some examples that may not fall into this general category (transformation by some of the oncogenic DNA viruses is one of them). We must be wary of the ever-present occupational risk of the cancer researcher: generalization. With this caveat, it is quite clear that cellular oncogenes can be switched on to constitutive activity by a variety of mechanisms. In the model of direct retroviral transformation, the oncogene is switched on because it has recombined with a powerful retroviral promoter. The resulting unit can transform, no matter where it integrates. This situation is mimicked by the successful DNA-mediated transformation with large pieces of tumour-derived DNA⁸⁶. In other situations, the cellular oncogene remains in place, but becomes activated due to events that occur in its immediate neighbourhood. The integration of a viral promoter is one such event, chromosomal translocation to a highly active region of the cellular genome is another. Experimentally, the occasional successful transformation by normal cell-derived small DNA fragments probably mimics the latter situation.

The identification of the oncogenes and the definition of their regulation and function in the normal cell are clearly among the most fascinating aspects of the field. Their high evolutionary conservation strongly suggests that they have essential roles in cell growth and regulation. This analysis has already erased much of the traditional conceptual and experimental barriers between chemical and viral carcinogenesis and will no doubt eventually unite them under the common umbrella of cell genetics.

My studies and those of my colleagues in this area have been supported by grant 2 R01 CA 14054-07A1, awarded by the National Cancer Institute, DHEW and by grant NC-6580 from the Swedish Cancer Society.

1. Cairns, J. *Nature* **289**, 353-357 (1981).
2. Bishop, J. M. *Cell* **23**, 5-6 (1981).
3. Stchelin, D. *et al. Cold Spring Harb. Symp. quant. Biol.* **44**, 1215-1223 (1980).
4. Eriksson, R. L. in *Viral Oncology* (ed. Klein, G.) 39-53 (Raven, New York, 1980).
5. Rosenberg, N. & Baltimore, D. in *Viral Oncology* (ed. Klein, G.) 187-203 (Raven, New York, 1980).
6. Sefton, B. M., Hunter, T., Beemon, K. & Eckhart, W. *Cell* **20**, 807-816 (1980).
7. Hunter, T. *Cell* **22**, 647-648 (1980).
8. Graf, T. & Beug, H. *Biochim. biophys. Acta* **516**, 269-278 (1978).
9. Graf, T., Ade, N. & Beug, H. *Nature* **275**, 496-501 (1978).
10. Hayward, W. S., Neel, B. G. & Astrin, S. M. *Nature* **290**, 475-480 (1981).
11. Oscarsson, M., McClements, W. L., Blair, D. G., Maizel, J. V. & van de Woude, G. F. *Science* **207**, 1222-1224 (1980).
12. Jones, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 2651-2655 (1980).
13. van Beveren, Ch. *et al. Nature* **289**, 258-262 (1981).
14. Foulds, I. M. A. *J. chronic Dis.* **8**, 2-37 (1958).
15. Martin, R. G. *Adv. Cancer Res.* **34**, 1-68 (1981).
16. Howley, P. M. in *Viral Oncology* (ed. Klein, G.) 489-550 (Raven, New York, 1980).
17. Graessmann, A., Graessmann, M. & Mueller, Ch. *Adv. Cancer Res.* (in the press).
18. Lane, D. P. & Crawford, L. V. *Nature* **278**, 261-263 (1979).
19. Melero, J. A., Stitt, D. T., Mangel, W. F. & Carroll, R. B. *Virology* **93**, 466-480 (1979).
20. Linzer, D. I. H. & Levine, A. J. *Cell* **17**, 43-52 (1979).
21. Crawford, L. V., Pim, D. C., Gurney, E. G., Godfellow, P. & Taylor-Papadimitriou, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 41-45 (1981).
22. Gurney, E. G., Harrison, R. O. & Fenno, J. J. *J. Virol.* **34**, 752-763 (1980).
23. Rotter, V., Witte, O. N., Coffman, R. & Baltimore, D. *J. Virol.* **36**, 547-555 (1980).
24. Linzer, D. I. H., Maltzman, W. & Levine, A. J. *J. Virol.* **98**, 308-318 (1979).
25. Kress, M., May, E., Cassingena, R. & May, P. J. *J. Virol.* **31**, 472-483 (1981).
26. DeLeo, A. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 2420-2424 (1979).
27. Dippold, W. G., Jay, G., DeLeo, A. B., Khoury, G. & Old, L. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1695-1699 (1981).
28. Reedman, B. M. & Klein, G. *Int. J. Cancer* **11**, 499-520 (1973).
29. Luka, J., Jörnvall, H. & Klein, G. *J. Virol.* **35**, 592-602 (1980).
30. Klein, G., Luka, J. & Zeuthen, J. *Cold Spring Harb. Symp. quant. Biol.* **44**, 253-261 (1980).
31. Graessmann, A. W. & Bornkam, G. W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 433-436 (1980).
32. Volsky, D. J., Shapiro, I. M. & Klein, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5453-5457 (1980).
33. Shapiro, I. M., Klein, G. & Volsky, D. *Biochim. biophys. Acta* **676**, 19-24 (1981).
34. Volsky, D. J., Klein, G., Volsky, B. & Shapiro, I. M. *Nature* **293**, 399-401 (1981).
35. Yefenof, E. & Klein, G. *Expl Cell Res.* **99**, 175-178 (1976).
36. Yefenof, E., Klein, G., Ben-Bassat, B. & Lundin, L. *Expl Cell Res.* **108**, 185-190 (1977).
37. Steinitz, M. & Klein, G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3518-3520 (1975).
38. Steinitz, M. & Klein, G. *Virology* **70**, 570-573 (1976).
39. Montagnier, L. & Gruet, J. *Int. J. Cancer* **23**, 71-75 (1979).
40. Ernberg, I. *et al.* (submitted).
41. Rigby, P. *Nature* **282**, 781-784 (1979).
42. Crawford, L. V. *et al. Cold Spring Harb. Symp. quant. Biol.* **44**, 179-188 (1980).
43. Ito, Y. in *Viral Oncology* (ed. Klein, G.) 447-480 (Raven, New York, 1980).
44. Benjamin, T. L. & Schaffhausen, B. S. *Cell* **18**, 935-946 (1979).
45. Soeda, E., Maruyama, T., Arrand, J. R. & Griffin, B. E. *Nature* **285**, 165-167 (1980).
46. Rowe, W. P. *J. exp. Med.* **136**, 1272-1285 (1972).
47. Wiener, F., Ohno, S., Spira, J., Haran-Ghera, N. & Klein, G. *J. natn. Cancer Inst.* **61**, 227-238 (1978).
48. Wiener, F., Spira, J., Ohno, S., Haran-Ghera, N. & Klein, G. *Int. J. Cancer* **22**, 447-453 (1978).
49. Wiener, F., Ohno, S., Spira, J., Haran-Ghera, N. & Klein, G. *Nature* **275**, 658-660 (1978).
50. Spira, J., Wiener, F., Ohno, S. & Klein, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6619-6621 (1979).
51. Wiener, F. *et al. Int. J. Cancer* **27**, 51-58 (1981).
52. Falkow, P. J., Reddy, A. L. & Bryant, J. I. *Int. J. Cancer* **26**, 603-608 (1980).
53. Spira, J. *et al. Leukaemia Res.* **5**, 113-121 (1981).
54. Chang, T. D., Biedler, J. L., Stockert, E. & Old, L. J. *Proc. Am. Ass. Cancer Res.* (abstr.), 225 (1977).
55. Wiener, F., Spira, J., Ohno, S., Haran-Ghera, N. & Klein, G. *Int. J. Cancer* **23**, 504-507 (1979).
56. Wiener, F., Spira, J., Babonits, M., Haran-Ghera, N. & Klein, G. *Int. J. Cancer* **26**, 661-668 (1980).
57. Dofuku, R., Riedler, J. L., Spengler, B. A. & Old, L. J. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1515-1517 (1975).
58. Spira, J. *et al. Cancer Res.* **40**, 2609-2616 (1980).
59. Lilly, F. J. *natn. Cancer Inst.* **45**, 163-169 (1970).
60. Lilly, F. & Pincus, T. *Adv. Cancer Res.* **17**, 231-277 (1973).
61. Kaplan, H. S. *Cancer Res.* **16**, 434-436 (1956).
62. Law, L. W. *Ann. N.Y. Acad. Sci.* **68**, 616-618 (1957).
63. Rowe, W. P. *Cold Spring Harb. Symp. quant. Biol.* **44**, 1265 (1979).
64. Spira, J., Wiener, F., Babonits, M., Miller, J. & Klein, G. *Int. J. Cancer* (in the press).
65. Ohno, S. *et al. Cell* **18**, 1001-1007 (1979).
66. Wiener, F., Fenyö, E. M., Klein, G. & Davies, A. J. S. *Somatic Cell Genet.* **2**, 81-92 (1976).
67. Meo, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 550-553 (1980).
68. Hengartner, H., Meo, T. & Müller, E. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4494-4498 (1978).
69. Wiener, F., Babonits, M., Spira, J., Klein, G. & Potter, P. *Somatic Cell Genet.* **6**, 731-738 (1980).
70. Hieter, P. A., Korsmeyer, S. J., Valdmann, T. A. & Leder, P. *Nature* **290**, 368-372 (1981).
71. Manolov, G. & Manolova, Y. *Nature* **237**, 33-34 (1972).
72. Zech, L., Haglund, U., Nilsson, K. & Klein, G. *Int. J. Cancer* **17**, 47-56 (1976).
73. Manolova, Y., Manolov, G., Kieler, J., Levan, A. & Klein, G. *Hereditas* **90**, 5-10 (1979).
74. Mitelman, F. *et al. Int. J. Cancer* **24**, 27-33 (1979).
75. Berger, R., Bernheim, A., Broet, J. C., Daniel, M. T. & Flandrin, G. *Br. J. Haemat.* **43**, 87-90 (1979).
76. Rowley, J. D. *Cancer Genet. Cytogenet.* **1**, 263-271 (1980).
77. Mitelman, F. *Adv. Cancer Res.* **34**, 141-170 (1981).
78. Tanzer, J., Frocrain, C., Alcalay, D. & Desmarest, M. C. *Nouv. Revue fr. Hémat. Suppl.* **22**, 107 (1980).
79. Miyoshi, I., Hiraki, S., Kimura, I., Miyamoto, K. & Sato, J. *Experientia* **35**, 742-743 (1979).
80. Van Den Berghe, H. *et al. Cancer Genet. Cytogenet.* **1**, 9-14 (1979).
81. Berger, R. *et al. Hum. Genet.* **53**, 111-112 (1979).
82. Rowley, J. D. *Cancer Genet. Cytogenet.* **2**, 175-198 (1980).
83. Croce, C. M. *et al. Eur. J. Immun.* **10**, 486-488 (1980).
84. McKusick, V. A., *Cytogenet. Cell Genet.* (in the press).
85. Cooper, G. M., Okenqvist, S. & Silverman, L. *Nature* **284**, 418-421 (1980).
86. Shih, C., Shilo, B.-Z., Goldfarb, M. P., Dannenberg, A. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5714-5718 (1981).
87. Rigby, P. W. J. *Nature* **290**, 186-187 (1981).

ARTICLES

New observational constraints on the M87 jet

J. T. Stocke*, G. H. Rieke*[†] & M. J. Lebofsky** Steward Observatory, and [†] Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

New observations at 1.6–3.45 μm confirm the presence of a dramatic ($\Delta\alpha \sim 1$) break between radio–IR wavelengths and 6,000 Å, in the spectrum of the M87 jet. These data, in combination with data taken in other spectral regions, show that the individual knots in the M87 jet have nearly the same spectral indices and nearly the same large ($\Delta\alpha \sim 1$) spectral break. This large spectral break and the constancy of spectral properties between the knots pose serious constraints for models of the M87 jet.

THE M87 jet has long attracted detailed observational and theoretical interest¹ because of its clear relationship with non-thermal activity centred in the nuclei of galaxies. Recently this object has become the prototype for a myriad of radio galaxies which possess radio (and in some instances optical and even X-ray) jets which clearly link the galactic nucleus to the extended regions of radio emission (for a review see ref. 2).

The M87 jet is unique in that it is easily detectable over a very wide range of frequency (10^8 – 10^{17} Hz) and, therefore, gives us the opportunity to study the distribution of photon energies in some detail. Thus, the overall spectrum of the M87 jet may help to determine the exact manner in which energy is transported from the nucleus to the outer radio emitting regions. We now present new IR observations of the brightest knots in the M87 jet at H (1.6 μm), K (2.2 μm) and L (3.4 μm).

New observations

Near IR measurements of the jet of M87 were obtained on the nights of 17 and 18 February 1981, using the Steward Observatory 2.25-m telescope. The seeing on these nights was generally good and in particular was 1 arc s or better during the H measurements. Three different observing modes were used. First at K (2.2 μm), where the flux from the jet dominates the observed flux, consisted of photometry with the aperture centred on knot A, 12 arc s from the nucleus and complementary photometry centred on a patch of galaxy 12 arc s from the nucleus but 180° different in position angle. In the second mode,

Table 1 IR photometry of knots A + B in the M87 jet

Wavelength (μm)	Flux (mJy)
H (1.6)	3.8 ± 0.2
K (2.2)	5.1 ± 0.2
L (3.45)	7.2 ± 0.6

at L (3.45 μm), where the stellar background makes only a small contribution, photometry on the jet was obtained, and the stellar contribution was estimated from the 2- μm flux of the complementary galaxy patch and assuming that $K - L = 0.2$. This correction amounted to only 10% of the flux observed at L . The third observing mode was at H (1.6 μm), where the stellar contribution is comparable with that from the jet, and consisted of scans in the jet and anti-jet directions, registered on the nucleus. The result of this scanning at 1.6 μm is shown in Fig. 1 where the flux from the jet is clearly seen as an excess above the stellar flux. A 5.9 arc s aperture, chopping in declination, and a reference beam displaced 10 arc s in declination from the signal beam were used for all the measurements. The observations were calibrated as described by Low and Rieke³; the zero point of the magnitude scale at H was assumed to be 1,080 Jy. Data at 2.2 and 3.45 μm were taken using a liquid-nitrogen cooled InSb detector while a liquid-helium cooled InSb detector was used at 1.6 μm and to repeat the observations at 2.2 μm . In addition, the stellar contribution at K was checked by estimating the stellar flux at 1.6 μm from the scan and assuming that $H - K = 0.25$, the stellar colour appropriate at knot A's distance from the nucleus. This procedure yielded a stellar correction within 20% of the value observed from the complementary galaxy patch. These results are listed in Table 1.

Comparison with observations at other wavelengths

Before these IR observations, there was still some question as to whether the radio emission from the jet arose from the same physical source as the optical emission. On the one hand recent interferometer maps of the jet by Owens *et al.*⁴ and by Laing⁵ strongly argue for a single emission mechanism as they demonstrate a one-to-one correspondence between the position, size and relative brightness of the radio and optical knots. However, there were lingering doubts about this conclusion because:

- (1) Earlier attempts⁶ to connect all the available optical, radio and IR detections and upper limits with a single smooth curve produced a violation or near violation of the 10 μm upper limit.
- (2) The optical spectral index of 1.7 is much larger than the radio spectral index of 0.6. The procedure of fitting a smooth curve through the data required a dramatic break ($\Delta\alpha \sim 1.0$) just below the lowest frequency optical observation⁷. Such a large

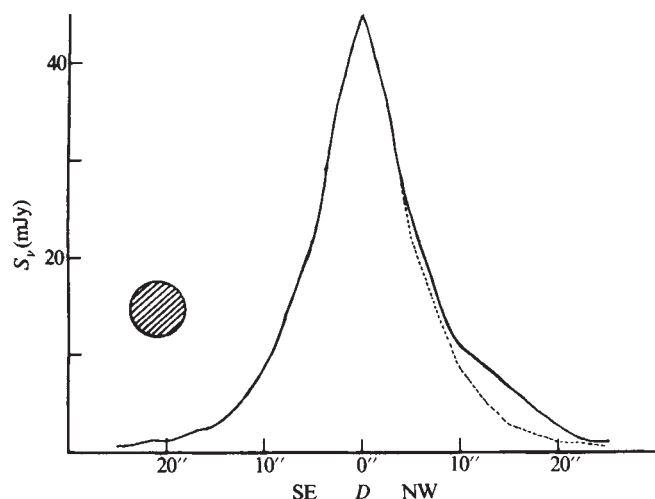


Fig. 1 Scan through the nucleus of M87 from the south-east to the north-west (jet) at H (1.6 μm) using a beam as indicated by the shaded circle. D is the distance from the nucleus. The dashed line on the north-west side is the reflection of the south-east (stars only) scan about the nucleus.

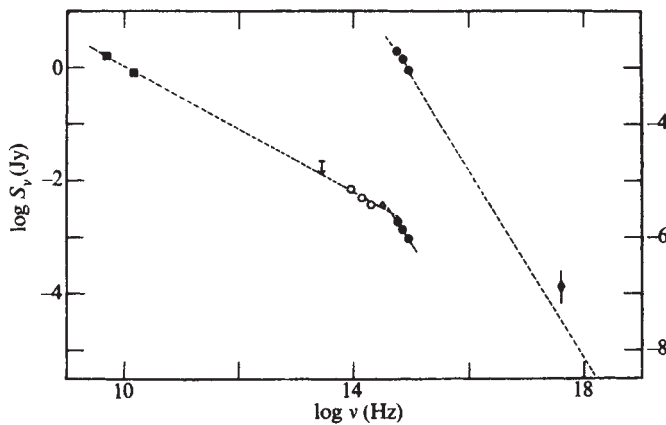


Fig. 2 The spectrum of the brightest knots ($A + B$) in the M87 jet. The radio ($\blacksquare$)⁴⁻⁵, IR ($\circ$, this paper, 10- μ m upper limit from ref. 6), optical ($\triangle$ for I (ref. 7); $\bullet$, for U, B, V (refs 6, 8, 9)), and X-ray ($\blacklozenge$)¹² have been normalized to a 6 arc s aperture centred on knot A , 12 arc s west and 5 arc s north of the nucleus. The X-ray point is a flux measurement of the entire jet corrected downwards by assuming a light distribution identical to that seen in the optical^{8,12}. The left-hand flux scale refers to radio-optical frequencies while the right-hand scale refers to optical-X-ray frequencies. The errors are smaller than the figure symbols except where explicitly shown.

break is twice as large as that expected due to isotropic synchrotron losses and continuous particle injection.

Figure 2 shows the spectrum of the region of knots A and B (see ref. 7 for knot identification) including our IR measurements at H, K and L . All of these observations are corrected to a 6 arc s aperture at the position of knot A , 12 arc s from the nucleus. At this position the aperture includes knots A, B, I , and J . The optical observations⁶ have been corrected downward by 0.03 mag at each frequency because knot K was included in the 7-arc s aperture used (m_B for knot $K = 18.95 \pm 0.2$)⁸. Only if the positioning of the observing aperture were displaced 1 arc s would knot C be included, requiring an additional 0.25 mag downward correction. That this is not the case is suggested by the agreement between the photoelectric blue magnitude⁶ and the photographically determined blue magnitudes⁸ of knots A, B, I and J . The point at I (ref. 7) does include knot C and has been corrected downward, but it is still 0.6 mag too high to fit a smooth curve between V and H . It is not surprising that this point is the most discrepant because it is based on photographic photometry near the wavelength where the contrast between the jet and the galaxy is minimal. Note that there is no indication of recent (1964–74) optical variability in knots A and B (± 0.1 mag)⁷⁻⁹ although there is a claim for optical variability on a longer time scale (1956–78)¹⁰.

The radio data are from ref. 5 at 15.4 GHz and ref. 4 at 5 and 15 GHz. At 15 GHz the values of the peak flux and knot size for knots A and B from ref. 5 yield total flux values 60% higher than similar estimates from the map in ref. 4. We have used the ref. 5 value both because the larger beam size may have included flux missed by the smaller beam of the incomplete VLA and because of the uncertainty in determining peak fluxes from contour maps. From the 805 ± 25 mJy value at U band⁵, we estimated $1,550 \pm 60$ mJy at C band by using the spectral index between those two bands ($\alpha = 0.59 \pm 0.08$)⁴.

We have not plotted the UV data points for the jet from observations made with IUE¹¹ because there was no attempt to subtract the galaxy's contribution accurately. Although these points are very uncertain, they do suggest a continuation of the optical power law from 3,000 to 1,000 Å without a dramatic change. A recent X-ray point¹² is also consistent with an extrapolated power law of $\alpha = 1.7$ to within the errors. There is also an indication that the distribution of light in X rays is similar to that in the optical. The high frequency data are thus consistent at present with a single steep power law from $\log \nu = 14.8$ to 17.6.

The new data for the jet are consistent with a smooth spectrum with internal slopes in the radio, IR and optical of 0.6, 0.8 and 1.7. The point at I (ref. 7) is not consistent with this spectrum, but the 10 μ m upper limit is no longer violated. If we demand that the extrapolation of the IR data points to higher frequencies fall above or at the point at V (ref. 6) then the IR slope must be 0.7 ± 0.1 . Thus, the data now argue strongly that the radio and IR emission mechanism is the same because the measurements in both spectral regions connect up with a simple smooth curve of nearly constant slope. In fact, the present observations are still consistent with a single power law below the optical with $\alpha = 0.7$, although there is a slightly better fit with a gently curving spectrum steepening from 0.55 to 0.7. Either case requires a dramatic break between IR and optical and further constrains the spectrum so that a break of $\Delta\alpha \sim 1$ most probably occurs between 5,500 and 6,500 Å and certainly within the range 4,500–8,000 Å. The slope from H (1.6 μ m) to V is 0.6, indicating that the break cannot be far from V .

The spectral break could be due to reddening, but this seems improbable because: (1) the IUE data¹¹ show no evidence for the 2,200 Å dust feature in the jet; (2) without any dereddening being applied, the V point lies slightly above the extrapolated power law from the IR—dereddening will increase this difference; (3) scans along the jet indicate spectral properties which are constant along the jet so the extinction along the jet would have to be uniform over many kiloparsecs; (4) internal reddening has not been observed in elliptical galaxies.

The scan in Fig. 1 can be used to derive the IR brightness along the jet. Figure 3 shows the results of a point-by-point subtraction in the form of a strip scan along the jet with 6 arc s resolution. The data (solid line) are compared with similar data at 15.4 GHz (ref. 5) and in the optical U filter⁹. A similar optical scan in the B filter¹³ is identical to within the combined 3% errors with the U filter data. We have plotted only the U filter data as its errors are slightly smaller due to the higher contrast of the jet at that frequency and to the use of a linear detector in that data set. No data are plotted inside 10 arc s from the nucleus where the galaxy subtraction causes much larger errors at optical and IR frequencies.

To within the combined errors all four scans are identical indicating extremely similar spectral indices all along the jet. Table 2 lists the radio-IR (low ν) and IR- U filter (high ν) spectral indices for the four brightest knots outside 10 arc s from the nucleus. To within quite small errors the spectra both above and below the break are identical for all four knots.

Even though no scan is available at V band, these data strongly suggest that the break frequency is identical for all four knots to within ± 0.3 in $\log \nu$.

New constraints on jet models

These new IR observations place severe restrictions on any successful model of the M87 jet.

- (1) The optical and radio emission almost certainly arise from a single emission mechanism.
- (2) The break in the spectrum of the brightest knots is very abrupt in frequency ($\Delta \log \nu \leq 0.1$) and large ($\Delta\alpha \approx 1$).
- (3) The spectra of the various knots are very similar both above and below the break frequency.
- (4) The break frequency must be nearly identical (to within 0.3 in $\log \nu$) for the brightest knots.

Table 2 Spectral indices in the M87 jet

Knot	α_{low}	α_{high}	λ_{break} (Å)
A	0.57 ± 0.01	0.92 ± 0.04	5,500–6,500
B	0.58 ± 0.01	0.93 ± 0.05	4,500–6,500
C	0.58 ± 0.01	0.92 ± 0.05	4,500–6,500
G	0.58 ± 0.02	0.88 ± 0.09	4,500–8,000

$$\alpha_{\text{low}} = \alpha(H (1.6 \mu\text{m}), 5 \text{ GHz}); \alpha_{\text{high}} = \alpha(U (0.36 \mu\text{m}), H (1.6 \mu\text{m})).$$

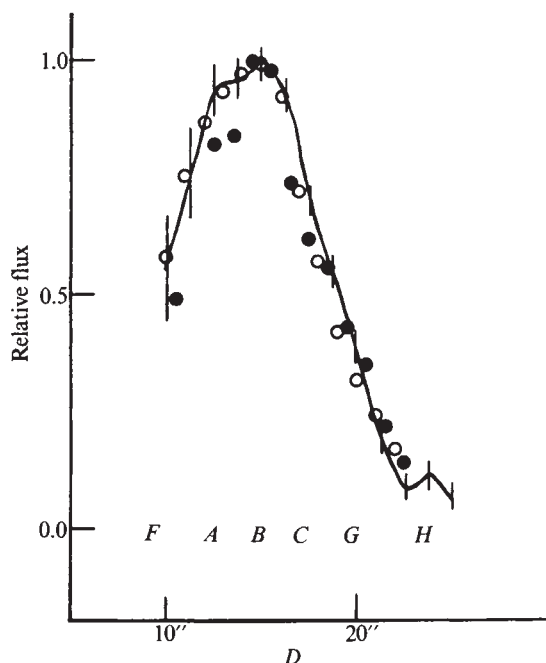


Fig. 3 Comparison of the flux distribution along the jet at three wavelengths: radio⁵ (○), IR from this paper (solid line), and optical^{9,13} (●). The radio and optical data have been convolved with the 6 arc s beam used for the IR measurements. The error bars refer to the IR data and are due mainly to uncertainties in subtracting the stellar background in M87. At the flux peak, the errors at all frequencies are ~3% and remain constant with D (distance from the nucleus) in the radio but increase towards the nucleus in the optical and IR. Beyond 20 arc s a flux of 38 mJy per beam has been subtracted from the radio points to eliminate the contribution from the extended ridge seen at 15 GHz (ref. 5). The letters refer to locations of the various knots⁸.

(5) The optical and near-IR emission dominate the observed energy losses from the jet. At a distance of 15.7 Mpc (ref. 14), the luminosity of knot A alone at V is 1.6×10^{41} erg s⁻¹ from a region less than 25 pc across⁸. The electrons responsible for this radiation have $\gamma \sim 10^6$ if we assume the equipartition value for the magnetic field strength ($H \sim 10^{-3}$ G)¹⁵.

In classical synchrotron theory¹⁶ a high frequency spectral break may be due to either (1) a break in the electron energy distribution or (2) energy dependent losses.

In regions of low radiation flux as in extended radio sources, synchrotron emission dominates high frequency energy losses. For an isotropic particle distribution and a continuous injection of electrons the expected break is $\Delta\alpha = 0.5$. A similar break is expected in regions of high radiation flux where Compton scattering dominates the energy losses. A more dramatic break is possible if either: (1) the particle pitch angle distribution is highly anisotropic or (2) the injection of electrons ceased a time t yr ago where:

$$t^2 = 4 \times 10^8 (H \sin \theta)^{-3} \nu_b^{-1} \quad (1)$$

where H is the magnetic field strength in G; θ the particle pitch angle; and ν_b the frequency of the spectral break.

The cessation of particle injection is an attractive hypothesis because the expected $\Delta\alpha$ matches the observations: $\alpha_{\text{low}} = (\gamma - 1)/2 = 0.6$ to $\alpha_{\text{high}} = (2\gamma + 1)/3 = 1.8$ (ref. 16). However, the time since cessation is extremely short compared with the light travel time between the knots or from the nucleus to the knots¹. For example, given an equipartition magnetic field strength of $H \geq 10^{-3}$ G, $t \leq 30$ yr (ref. 15). The nearly identical value that we find for the break frequencies in the four brightest knots requires that the time since last particle injection into each knot is identical to within 40% of the light travel time from the nucleus to knot A, assuming a magnetic field of $10^{-4.5}$ G. For larger fields this restriction is, of course, much worse. Therefore, an additional difficulty for the synchrotron loss model for the

spectral break is that no matter how the energy is supplied (for example, electrons or protons streaming from the nucleus or *in situ* acceleration), and given reasonable magnetic field values ($H \leq 10^{-4.5}$ G), the 'shut off' time for energy input is identical from knot-to-knot to less than the light travel time from knot-to-knot. Only a favourable projection angle can save this mechanism for consideration, a possibility that could be tested by further observations of the jet in the range 4,000–8,000 Å.

The expected spectral break is also steeper in the case of a highly anisotropic particle pitch angle distribution (or distribution in $H \sin \theta$ to be more precise). For electrons at a single pitch angle an exponential cutoff is observed at high frequencies even for continuous particle injection. This is consistent with the optical and UV observations¹¹ which approximate a very steep power law between 1,000 and 5,000 Å but could exponentially decay in the unobservable EUV. However, if the observed X-ray point is an extension of the optical-UV power law, the spread in pitch angle (assuming a constant magnetic field strength) must be at least an order of magnitude in $\sin \theta$ with more electrons populating the larger pitch angles and thus responsible for the observed optical radiation. Either of these two situations is possible if particles streaming at low pitch angles (and thus with long synchrotron lifetimes) are pitch-angle-scattered when they reach the observed knots. However, the great similarity among the high frequency spectral indices of the individual knots means that the scattering process must be so uniform in knots separated by thousands of light years that the particle distributions in $H \sin \theta$ are virtually identical from knot to knot. Perhaps knot-to-knot variations do exist in the spectra above the break but the present observations are not accurate enough to reveal them.

Based on the above, we favour interpreting the shape of the spectrum of the jet as being due to the underlying electron energy distribution. If this is true it is probably then that the X-ray flux arises from Compton scattering. Otherwise, to avoid an additional high frequency spectral break, the radiating particles should show negligible synchrotron losses even for the $\gamma \sim 10^7$ electrons responsible for the X rays. The very short synchrotron loss time scale for these electrons (1–200 yr for $H = 10^{-3}$ – $10^{-4.5}$ G) limits the size of the emitting region. An equipartition field strength of $H \geq 10^{-3}$ G would require that each knot be smaller than 4 m arc s. The expected Compton-scattered X-ray flux would then exceed the observed flux by about an order of magnitude¹⁷. This discrepancy can be reduced but not entirely removed by assuming a much smaller magnetic field, or by requiring the acceleration to take place along a line or in a very thin slab as might be expected with a shock front acceleration model. Thus, the X-ray radiation can be of synchrotron origin only if anisotropies in the radiation field and the electron velocity distribution reduce the Compton scattering substantially.

The inverse Compton origin of the X-ray radiation can be tested by observing the overall jet spectrum in the Einstein band. A second frequency observation would be expected to detect the difference between the expected slopes for a synchrotron ($\alpha = 1.7$) and inverse Compton ($\alpha \leq 0.8$) origin for the X rays. The inverse Compton X-ray slope is expected to be the same as the observed low frequency synchrotron slope ($\nu \leq 10^{14}$).

If the photon spectrum does reflect the underlying electron particle distribution, there are two possibilities for the origin of these electrons:

First, *in situ* acceleration overcomes the difficulties due to synchrotron losses described above but whatever acceleration process is responsible for the high energy electrons, it must be remarkably uniform in sites separated by thousands of light years in M87. As in the particle pitch angle scattering model for the high frequency break, the expectation for *in situ* acceleration models is that the different knots will have different spectra reflecting variations in acceleration parameters (for example, magnetic field strength and length scale over which the acceleration process takes place).

The large characteristic γ s of 10^6 in the jet also pose a problem for statistical acceleration mechanisms involving magnetic fields. Typically $\gamma \sim H^2 dr/c$ for these types of processes (for example, Fermi or betatron acceleration where H^2/c is the magnetic field energy and dr is the length scale over which it is applied). Thus, for an equipartition field and knot sizes between 2 and 20 pc, the expected γ s are only 10^{2-3} . However, large γ s are expected for processes whereby electrons gain energy due to motion through an electric field such as could be produced at a shock front moving through a region containing a magnetic field¹⁸. For this class of models $\gamma \sim qH dr/c$ (q is the electron charge) and thus gives characteristic γ s of 10^{6-7} for the above parameters. This electric field acceleration model is also much less sensitive to variations in source size as $\gamma(r) \propto r^{-1}$ instead of r^{-3} for the magnetic acceleration models. However, in either case knot-to-knot variations would still be expected and have not yet clearly been seen.

Second, an external source of particle energy is possible if a proton beam is emanating from the nucleus¹. This removes the problem of synchrotron losses and naturally explains the identical spectra of all the knots. Such a proton beam produces electrons by p-p collisions and subsequent muon decay into energetic electrons and γ rays. Using the formalism of Gould and Burbidge¹⁹, Felten¹, derives the relationship between the γ -ray flux $\phi(>E)$ (photons $\text{cm}^{-2} \text{s}^{-1}$) above an energy E (eV) and the observed synchrotron spectrum $F(\nu_e)$ in $\text{W m}^{-2} \text{Hz}^{-1}$ with slope s where $\nu_e = 10^{-6} HE^2$:

$$\phi(>E) = 7 \times 10^8 [a(2s+1)]^{-1} (0.2)^s HE F(\nu_e) \quad (2)$$

where H is the magnetic field in G and $a(s)$ is a function of spectral index s (see ref. 20, p. 68). As the dominant energy electrons occur at the break frequency $\log \nu_e = 14.75$, for an equipartition field of $H = 10^{-3}$ G the γ -ray energies associated with break frequency photons are $E = 8 \times 10^{11}$ eV. These are γ s in the range to which Cerenkov light detectors are sensitive. Equation (2) predicts γ -ray fluxes of 10^{-11} photons $\text{cm}^{-2} \text{s}^{-1}$, an order of magnitude less than the present upper limit on the γ -ray flux from M87 above 10^{11} eV (ref. 21).

If we assume that the magnetic field strength is the equipartition value, then there is a relationship between the observed

angular size of the knots and the γ -ray for the M87 jet in the proton beam model:

$$\phi(>10^{11} \text{ eV}) \approx 1 \times 10^{19} \theta^{-6/7} \text{ d exp}(-28.8 + 0.55 \log \theta) \quad (3)$$

where θ is the angular size in m arc s. Unfortunately equation (3) predicts only an order of magnitude increase in ϕ between $\theta = 300$ and 1 m arc s . The γ -ray flux decreases almost as fast as the magnetic field strength decreases away from equipartition; therefore, the proton beam mechanism is easily consistent with present observation.

We conclude that there are four observations which will constrain theoretical models of the M87 jet even more than the present observations:

(1) Flux measurements for each knot in the jet between 4,000 and 8,000 Å to determine more accurately the high frequency spectral index and break frequency for the knots. These observations are crucial in determining whether an *in situ* process (acceleration or pitch angle scattering) is operating in the jet. These observations require a two-dimensional linear detector to subtract the contribution from the galaxy, which is quite strong at these wavelengths.

(2) Speckle interferometry and VLBI observations of the knots to determine more accurately the angular size of the knots and thus the equipartition magnetic field strength and inverse Compton fluxes. The optical size is crucial because of longer synchrotron lifetimes, the knot sizes may be larger in the radio than in the optical because of particle diffusion away from the acceleration region.

(3) Improved γ -ray observations at high energies, to constrain the proton beam model.

(4) A second frequency X-ray observation at HRI resolution (≤ 3 arc s) to determine the X-ray spectral index and thus whether the X-ray radiation is due to a high energy tail of the optical synchrotron emission or to inverse Compton scattering from lower energy ($\nu \leq 10^{14}$ Hz) photons.

We thank B. Binkert for help at the 2.25-m telescope, Ethan Schreier and Eric Feigelson for communicating X-ray results before publication, and John Scott and John Cocke for helpful discussions. This research was supported by the NSF and NASA.

Received 13 July; accepted 29 September 1981.

1. Felten, J. E. *Astrophys. J.* **151**, 861–879 (1968).
2. Miley, G. K. *Rev. Astr. Astrophys.* **18**, 165–218 (1980).
3. Low, F. J. & Rieke, G. H. *Methods of Experimental Physics* Vol. 12, Pt A, (ed. Carlton, N.) 415–462 (Academic, New York, 1974).
4. Owen, F. N., Hardee, P. E. & Bignell, R. C. *Astrophys. J. Lett.* **239**, L11–L15 (1980).
5. Laing, R. A. *Mon. Not. R. astr. Soc.* **193**, 427–437 (1980).
6. Kinman, T. D., Grasdale, G. L. & Rieke, G. H. *Astrophys. J. Lett.* **194**, L1–L4 (1974).
7. Kaneko, N., Nishimura, M. & Toyama, K. *Publ. astr. Soc. Jap.* **25**, 175–180 (1973).
8. de Vaucouleurs, G. & Nieto, J.-L. *Astrophys. J.* **231**, 364–371 (1979).
9. Ables, H. D. & Kron, G. E. *Astrophys. J.* **181**, 19–25 (1973).
10. Sulentic, J. W., Arp, H. C. & Lorre, J. *Astrophys. J.* **233**, 44–55 (1979).

11. Perola, C. & Tarengi, M. *Astr. Astrophys.* (in the press).
12. Schreier, E. J., Feigelson, E. D. & Gorenstein, P. X. (in preparation, 1981).
13. de Vaucouleurs, G., Angione, R. & Fraser, C. W. *Astrophys. Lett.* **2**, 141–146 (1968).
14. Mould, J., Aaronson, M. & Huchra, J. *Astrophys. J.* **238**, 458–470 (1970).
15. Turland, B. D. *Mon. Not. astr. Soc.* **170**, 281–294 (1975).
16. Kardashev, N. S. *Soviet Astr. -AJ* **6**, 317–327 (1962).
17. Rieke, G. H. & Weekes, T. C. *Astrophys. J.* **155**, 429–437 (1969).
18. Pesses, M. thesis, Univ. Iowa (1979).
19. Gould, R. J. & Burbidge, G. R. in *Hdbuch Phys* **46/2**, 265–309 (1967).
20. Ginzberg, V. L. & Syrovetskij, S. I. *The Origin of Cosmic Rays* (transl. Massey, H. S. H.) (Pergamon, Oxford 1964).
21. Weekes, T. C., Fazio, G. G., Helmken, H. F., O'Mongain, E. & Rieke, G. H. *Astrophys. J.* **174**, 165–179 (1972).

Sm–Nd age of Kambalda and Kanowna greenstones and heterogeneity in the Archaean mantle

M. T. McCulloch & W. Compston

Research School of Earth Sciences, Australian National University, Canberra, ACT 2600, Australia

A precise Sm–Nd age of $2,790 \pm 30$ Myr and an initial ratio corresponding to $\epsilon_{\text{Nd}} = +3.4 \pm 0.6$ has been determined for the Kambalda sequence from the eastern Yilgarn Block of Western Australia. The high initial $^{143}\text{Nd}/^{144}\text{Nd}$ implies a long-lived light rare-earth element-depleted (high Sm/Nd) source for the Kambalda volcanics and therefore provides definitive evidence for ancient heterogeneities in the Archaean mantle.

THE Sm–Nd isotopic systematics have been examined in mafic–ultramafic and felsic volcanics at Kambalda and Kanowna in the Yilgarn Block of Western Australia. This study aims (1) to determine the original ages of the volcanic complexes, which have been difficult to establish by other methods, and (2) to determine their initial $^{143}\text{Nd}/^{144}\text{Nd}$ as a means of further understanding the evolutionary history of the crust and mantle. The Kambalda and Kanowna volcanics are typical of the greenstone belts which occur within the gneissic and granitic rocks of the Eastern Goldfields Province and have similarities with Archaean greenstones from other cratons. For example, the common occurrence of spinifex-textured Mg-rich basalts and peridotites of komatiitic affinity are analogous to those found in the Barberton area of southern Africa and Munro Township, Canada^{1–4}.

Despite intensive isotopic studies, the original ages of extrusion for the greenstones in this part of the East Yilgarn Block remain equivocal. Rb–Sr and Pb–Pb age determinations^{5–8} register specific events within the range ~2,760–2,400 Myr. However, because of the extensive metamorphic and metasomatic history of the terrain, even the oldest of these is interpreted as a minimum for the original igneous ages. Recent applications of the Sm–Nd method to dating of greenstones from other Archaean terrains^{9–12} have shown rare earth elements (REE) are much less mobile during metamorphism, so that the Sm–Nd system may be more conclusive for the definition of original ages. Due to the lack of preservation of primary mineralogy, the present study is based on whole rock analyses.

Volcanics have been analysed from two localities (see Fig. 1), Kambalda which is composed mainly of mafic and ultramafic extrusives and Kanowna where felsic volcanism predominates. Tentative stratigraphic correlations^{13,14} suggest that Kanowna is older, being part of a first volcanic cycle while Kambalda is thought to be at the base of a second volcanic cycle.

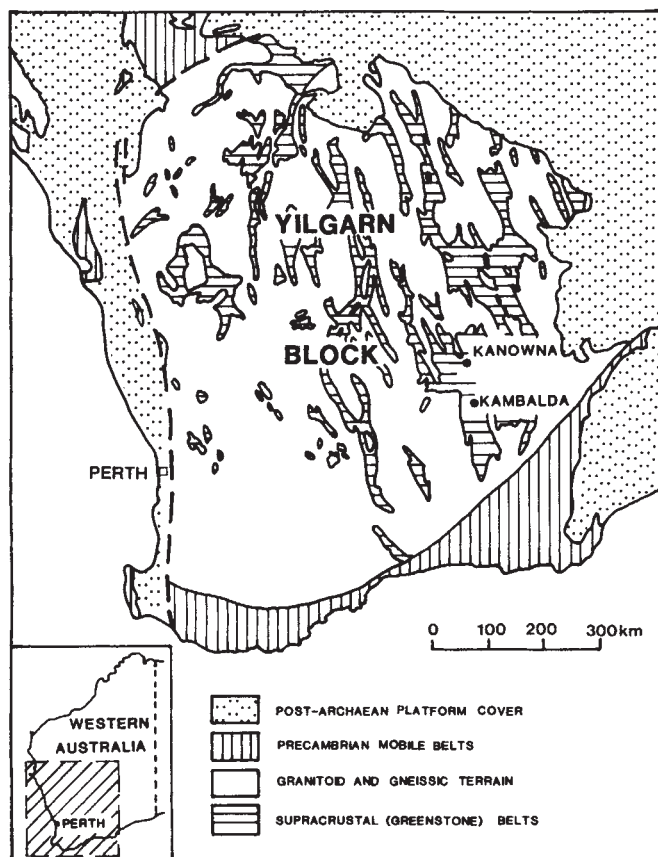


Fig. 1 Geological map of the Yilgarn Block of Western Australia showing locations of the Kambalda and Kanowna greenstone sequences.

Table 1 Sm–Nd isotopic data

Sample		Sm (p.p.m.)	Nd	$^{147}\text{Sm}/^{144}\text{Nd}^*$	$^{143}\text{Nd}/^{144}\text{Nd}^+$
Kambalda					
Footwall	72-5	1.83	5.51	0.2005	0.512094 ± 24
	72-13	1.54	4.80	0.1941	0.511948 ± 30
	72-19	2.10	6.52	0.1948	0.511954 ± 20
Hanging wall	72-916	1.97	6.60	0.1809	0.511716 ± 18
	72-910	1.78	6.04	0.1783	0.511680 ± 32
Ultramafic	KR3	0.79	2.14	0.2234	0.512537 ± 24
	72-42b	1.03	5.20	0.1198	0.510449 ± 18
		1.03	5.22	0.1197	0.510409 ± 26
Na-granite	71-751	1.91	12.55	0.0920	0.510088 ± 20
	71-752	2.16	13.15	0.0995	0.510226 ± 18
Felsic porphyry	72-41	2.11	13.52	0.0946	0.510131 ± 32
Kanowna					
Basalt	80-184	2.78	11.39	0.1477	0.511211 ± 18
	80-185	2.46	10.04	0.1481	0.511122 ± 22
Ultramafic	80-186	0.41	1.32	0.1896	0.512065 ± 20
Felsic porphyry	80-187a	3.12	18.49	0.1019	0.510280 ± 26
	80-187b	4.23	25.59	0.0999	0.510247 ± 18
Dacite Tuff	80-188	4.10	23.81	0.1043	0.510196 ± 18

* Uncertainties ±0.1%. Tracer solutions of ^{150}Nd and ^{147}Sm were calibrated using mixed normal solutions prepared from ultrapure pieces of Nd and Sm metal (Ames Laboratory) as gravimetric standards.

† Measured on spiked samples; normalized to $^{146}\text{Nd}/^{142}\text{Nd} = 0.636151$. The value for BCR-1 determined in this laboratory is 0.51184 ± 2 .

The Kambalda volcanics form a conformable succession consisting of an ultramafic unit some 240–600 m thick sandwiched between two basaltic sequences¹⁵. The lower basalt (the 'footwall basalt') is dominantly a massive metatholeiite with pillowed and flow-top breccia horizons and has many geochemical affinities with modern-day oceanic tholeiites⁴. The upper 'hanging wall' basalt is more heterogeneous and contains a lower variolitic-textured section separated by a sulphide rich sedimentary horizon from an upper massive fine-grained metatholeiite. The Mg-rich ultramafic sequence consists of a large number of flow units many of which have spinifex textures. Major nickel sulphide deposits occur at the base of the ultramafic sequence. Intruding the mafic–ultramafic sequence is a series of felsic intrusives of which a large portion consists of a 'sodic-granite'¹⁵. It has been suggested¹⁶ that these intrusives are recrystallized porphyries representing the subvolcanic expression of felsic volcanism within the mafic–ultramafic succession. The Kanowna area consists of a major felsic volcanic complex which overlies a sequence of metabasalts and ultramafics¹³. The metabasalts consist of a sequence of pillowed, amygdaloidal basaltic flows within which occurs several meta-peridotites. These rocks are thought to be the stratigraphically lowest unit in the Kalgoorlie area. The overlying felsic volcanics consist of lapilli tuffs, pyroclastic flows, agglomerates and a conglomerate containing felsic and ultramafic clasts in a predominately felsic volcanogenic matrix. Intruding this sequence is an altered quartz-rich porphyry. Gold has been produced from a quartz reef in the conglomerate.

Results

The procedures used for the isotopic analyses of Sm and Nd are described elsewhere¹⁷. In this study, samples were totally spiked and dissolved using HF, HClO_4 and HCl. Measurements were made of Nd⁺ using triple filaments. The total Nd blank is 100 pg and is therefore negligible for our study. The ϵ_{Nd} values are defined as:

$$\epsilon_{\text{Nd}} = \left[\frac{(^{143}\text{Nd}/^{144}\text{Nd})_{\text{init}}^T}{(^{143}\text{Nd}/^{144}\text{Nd})_{\text{CHUR}}^T} - 1 \right] \times 10^4$$

where $(^{143}\text{Nd}/^{144}\text{Nd})_{\text{CHUR}}^T = (^{143}\text{Nd}/^{144}\text{Nd})_{\text{CHUR}}^0 - (^{147}\text{Sm}/^{144}\text{Nd})_{\text{CHUR}}^0 (e^{\lambda_{\text{Sm}} T} - 1)$. Present-day reference values²⁶ for the chondritic unfractionated reservoir (CHUR)¹⁸ are $(^{143}\text{Nd}/^{144}\text{Nd})_{\text{CHUR}}^0 = 0.511836$ and $(^{147}\text{Sm}/^{144}\text{Nd})_{\text{CHUR}}^0 = 0.1967$; $\lambda_{\text{Sm}} = 6.54 \times 10^{-12} \text{ yr}^{-1}$. $(^{143}\text{Nd}/^{144}\text{Nd})_{\text{init}}^T$ is the measured ratio in the rock, corrected for decay since the time of

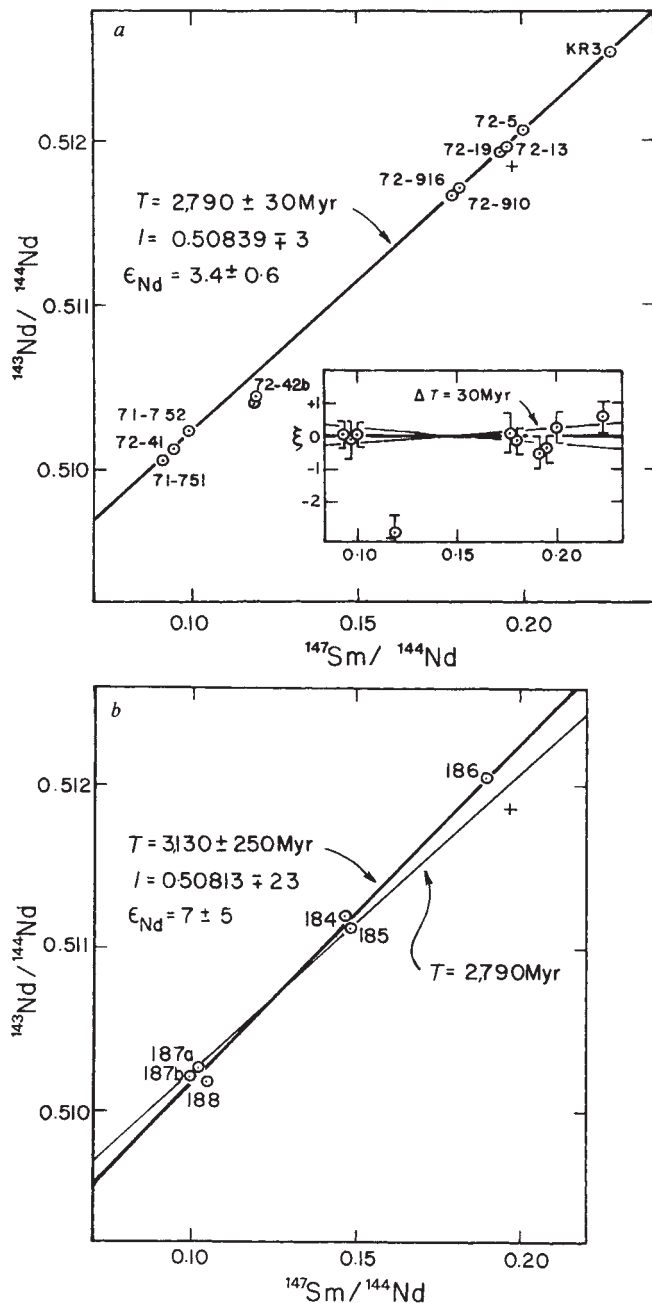


Fig. 2 *a*, Sm-Nd evolution diagram for the Kambalda volcanic sequence. The precise age and initial $^{143}\text{Nd}/^{144}\text{Nd}$ result from combining the basalts and ultramafics with an associated sodic granite and felsic porphyry. A less precise age of $2,910 \pm 170$ Myr is obtained from the footwall, hanging wall and ultramafics volcanics alone. +, The present-day chondritic value²⁶. *b*, Sm-Nd evolution diagram for the Kanowna volcanics. The large deviations from a best fit line do not allow a precise age determination.

crystallization T . The $^{147}\text{Sm}/^{144}\text{Nd}$ enrichment factor relative to 'CHUR' is given by

$$f_{\text{Sm/Nd}} = \left[\frac{(^{147}\text{Sm}/^{144}\text{Nd})_{\text{meas.}}}{(^{147}\text{Sm}/^{144}\text{Nd})_{\text{CHUR}}} - 1 \right]$$

The results of the isotopic measurements are listed in Table 1 and shown in Fig. 2*a, b*. Figure 2*a* shows the whole rock results from Kambalda. The footwall, hanging wall and ultramafic sequences from Kambalda have a relatively restricted range in Sm/Nd of $\sim \pm 10\%$ and approximately chondritic relative abundances of REE (Fig. 3). In contrast, the sodic-granite and associated porphyries have pronounced light rare-earth element (LREE) enrichments (low Sm/Nd). Combining all the

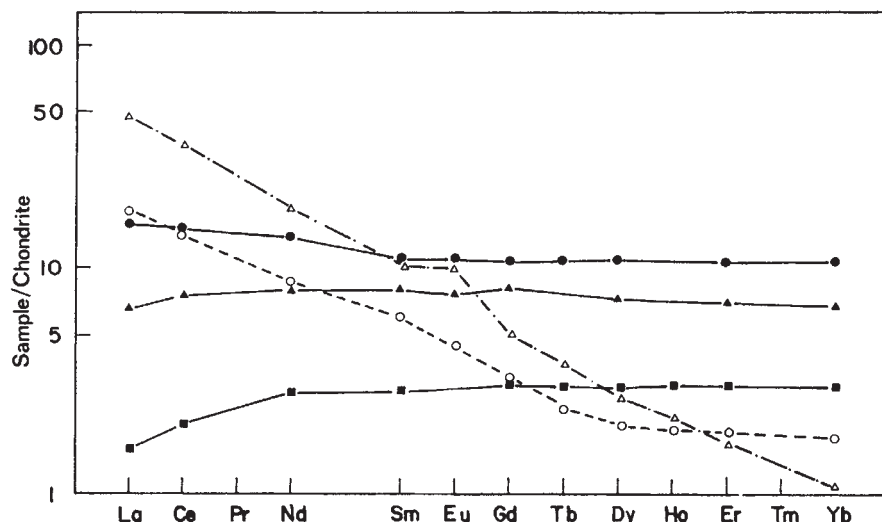
Kambalda samples yields a large range of Sm/Nd and therefore a large range in $^{143}\text{Nd}/^{144}\text{Nd}$ of 48 ϵ_{Nd} units. As shown inset in Fig. 2*a*, all the data points apart from an altered ultramafic lie on the best fit line within analytical uncertainty, and define an age of $2,790 \pm 30$ Myr (2σ), an initial $^{143}\text{Nd}/^{144}\text{Nd} = 0.50839 \pm 3$ (2σ). If it is assumed that the sodic-granite and felsic porphyry are isochronous with and have the same initial $^{143}\text{Nd}/^{144}\text{Nd}$ as the footwall, hanging wall and ultramafic volcanics, then the Sm-Nd age may represent the time of crystallization of this complex. This joint age is the same within errors as previous Pb-Pb (ref. 8) age determination for the sodic-granite of $2,760 \pm 70$ Myr. The strict co-linearity of the felsic and ultramafic data supports this contention. A less precise age of $2,910 \pm 170$ Myr is obtained from the footwall, hanging wall and ultramafic volcanics alone. The lower precision of ± 170 Myr is due to their smaller range in Sm/Nd. An ultramafic from the alteration zone at the felsic porphyry-ultramafic contact is the only discrepant sample. One explanation is that the Sm-Nd isotopic system in this sample has been disturbed by contamination with non-radiogenic Nd during a late alteration process associated with the emplacement of the felsic porphyry. The unusual LREE enrichment and abundances in the ultramafic (Fig. 3) is consistent with this interpretation. Alternatively the age obtained by combining the footwall, hanging wall and ultramafic volcanics with the altered ultramafic ($3,100 \pm 100$ Myr) could be registering a time of early alteration. However, this is unlikely due to the similarity between the LREE enrichments observed in the altered ultramafic and the younger felsic porphyry. The initial $^{143}\text{Nd}/^{144}\text{Nd}$ obtained from the $2,790 \pm 30$ Myr isochron lies above the CHUR evolution curve ($\epsilon_{\text{Nd}} = 3.4 \pm 0.6$).

Figure 2*b* shows the results from Kanowna. The samples show similar variation of Sm/Nd with bulk composition, the ultramafic having the highest Sm/Nd, metabasalts and intermediate Sm/Nd and the felsic samples the lowest Sm/Nd. Despite a range in $^{143}\text{Nd}/^{144}\text{Nd}$ of 36 ϵ_{Nd} units, the Kanowna samples produce a very imprecise age of $3,130 \pm 250$ Myr and initial $^{143}\text{Nd}/^{144}\text{Nd} = 0.50813 \pm 23$ ($\epsilon_{\text{Nd}} = 7 \pm 5$). For comparison, the Kambalda isochron is also shown in Fig. 2*b*. The large deviations from the best fit line of the Kanowna data are in distinct contrast with the Kambalda data and could be due to significant differences in age, initial $^{143}\text{Nd}/^{144}\text{Nd}$ or redistribution of Sm and Nd during later metamorphisms. In contrast to the Kambalda samples which are from deep drill-cores, the Kanowna samples are from the surface outcrops so we cannot exclude the possibility that their Sm-Nd isotopic system has been disturbed by recent weathering. As far as can be ascertained, there is no obvious difference in the extent of modification of equivalent rock types at Kanowna and Kambalda. For example, metabasalts from Kanowna and Kambalda have similar major element chemical contents (for example $\text{K}_2\text{O} < 0.1\%$) and ocelli bearing samples have been analysed⁷ from both localities (80-184 and 72-910, 72-916). Although disturbance of the Sm-Nd system by later metamorphic and metasomatic events cannot be discounted (for example, sample 72-42b) the consequences of age and initial $^{143}\text{Nd}/^{144}\text{Nd}$ variations will now be considered.

Sm-Nd whole rock isochrons

The evolution of initial $^{143}\text{Nd}/^{144}\text{Nd}$ with time in terrestrial rocks of varying ages has been shown^{9-12,18,24-26} to approximate the evolution expected in a reservoir having a fixed, chondritic Sm/Nd. For this reason, rocks which are closed systems and derived from the reference mantle reservoir (CHUR) will produce isochrons that pivot about the present-day values of Sm/Nd and $^{143}\text{Nd}/^{144}\text{Nd}$ in CHUR. This is shown in Fig. 4 for isochrons of age 1,000, 2,000 and 3,000 Myr. For example, a suite of rocks derived from a chondritic reservoir 3,000 Myr ago will today produce a linear array with the $^{143}\text{Nd}/^{144}\text{Nd}$ being proportional to Sm/Nd. A sample that has the same Sm/Nd as in the chondritic reservoir will continue to evolve in a manner identical to the reference reservoir. This applies to rocks of all ages and thus the present-day values of Sm/Nd and $^{143}\text{Nd}/^{144}\text{Nd}$

Fig. 3 Chondrite normalized REE abundance patterns for the Kambalda volcanics^{27,33}, sodic-granite²⁸ and altered ultramafic. The unusual LREE enrichment of the altered ultramafic may be due to contamination by LREE from the younger sodic-granite. Δ , Na-granite; $\bullet$, hanging wall basalt; $\circ$, altered ultramafic; $\blacktriangle$, footwall basalt; $\blacksquare$, ultramafic.



in CHUR will be a fixed point common to all isochrons derived from CHUR. Thus samples with Sm/Nd approximately the same as in CHUR ($f^{\text{Sm/Nd}} \approx 0$) will not provide meaningful age information. Samples which are most sensitive to age variations are those with non-chondritic Sm/Nd ($f^{\text{Sm/Nd}} \neq 0$). These systematics have important implications for this study (and others) as it has been necessary to combine felsic volcanics ($f^{\text{Sm/Nd}} \sim -0.5$) with mafic and ultramafic volcanics ($f^{\text{Sm/Nd}} \sim 0$) to obtain an adequate range in Sm/Nd for precise dating. At Kambalda the sodic granite and felsic porphyry (the low Sm/Nd end-members) intrude the mafic-ultramafic sequence and therefore the $2,790 \pm 30$ Myr 'isochron' represents only a minimum age for the mafic-ultramafic volcanics. It is possible, within the analytical uncertainties, that a significant age difference of 120 ± 200 Myr may exist between the mafic-ultramafic volcanics and felsic intrusives.

The disparate Kanowna data clearly imply disturbance of the Sm/Nd isotopic system or age and initial $^{143}\text{Nd}/^{144}\text{Nd}$ variations. For example, the large deviations between the low Sm/Nd felsic rocks requires that the felsic porphyry 187 has either a higher initial $^{143}\text{Nd}/^{144}\text{Nd}$ or is younger than the dacite 188. The geological field relationships are consistent with the latter explanation. The basalts 184 and 185 are spatially closely related suggesting that their deviations are more probably due to variable initial $^{143}\text{Nd}/^{144}\text{Nd}$. The contrast between the Kambalda and Kanowna data may be due to dominance of felsic volcanism at Kanowna. Felsic volcanics are chemically more evolved, implying isotopically more evolved and heterogeneous mantle or crustal sources.

The Sm-Nd model age^{18,19} ($T_{\text{CHUR}}^{\text{Nd}}$) is the time in the past when the major REE fractionation in a rock occurred during its derivation from the model mantle reservoir CHUR. In Figs 2 and 4 this is equivalent to the age defined by the line joining the sample with the present day values of $^{143}\text{Nd}/^{144}\text{Nd}$ and Sm/Nd in the chondritic reservoir CHUR. Thus only fractionated (non-chondritic Sm/Nd) samples define precise model ages. The $T_{\text{CHUR}}^{\text{Nd}}$ model ages from the low Sm/Nd Kambalda felsic volcanics, 2,510–2,540 Myr, are the same within analytical uncertainty whereas the Kanowna felsic volcanics have a range in $T_{\text{CHUR}}^{\text{Nd}}$ of 2,490–2,690 Myr. All these model ages are significantly younger than the ages indicated by the whole-rock isochrons. This discrepancy is due to the higher initial $^{143}\text{Nd}/^{144}\text{Nd}$ of the whole-rock isochrons compared with that in CHUR. This can be seen in Fig. 2a, b where the best fit lines to the mafic-ultramafic samples have a higher $^{143}\text{Nd}/^{144}\text{Nd}$ relative to CHUR for a chondritic Sm/Nd. The higher initial $^{143}\text{Nd}/^{144}\text{Nd}$ relative to CHUR (corresponding to $\epsilon_{\text{Nd}} = 3.4 \pm 0.6$ for Kambalda and $\epsilon_{\text{Nd}} = 7 \pm 5$ for Kanowna) is essentially independent of age for the mafic-ultramafic samples. The possible explanations for the well defined Kambalda point are discussed below.

Heterogeneous Archaean mantle

The well-defined initial $^{143}\text{Nd}/^{144}\text{Nd}$ of the Kambalda volcanics corresponding to $\epsilon_{\text{Nd}} = 3.4 \pm 0.6$ provides some of the first definitive evidence for ancient long-lived heterogeneities in the Archaean mantle. This can be seen in Fig. 5a where the initial $^{143}\text{Nd}/^{144}\text{Nd}$ determined on Archaean igneous rocks are summarized on a graph of ϵ_{Nd} versus time. Most of the previously determined 2,600–2,800 Myr data points cluster near the $\epsilon_{\text{Nd}}(T) = 0$ line suggesting that the Archaean mantle was relatively undifferentiated and the Sm/Nd of the Earth was uniform (within about $\pm 3\%$) and close to the average chondritic value used to calculate the CHUR curve²⁶. The Kambalda point (solid) lies distinctly above the CHUR curve and unambiguously implies a mantle source with $f^{\text{Sm/Nd}} > 0$ for time scales > 100 Myr. Its position above the CHUR curve is significant as the interpretation of data which lies below the CHUR curve is complicated by the possibility that the magmas may be contaminated with pre-existing continental crustal rocks with negative $\epsilon_{\text{Nd}}(T)$ values. This is exemplified by the Stillwater point²⁰. Evidence for the derivation of magmas from positive ϵ_{Nd} mantle has been presented for Archaean gneisses from eastern India by Basu *et al.*³¹ and for Proterozoic rocks by DePaolo³⁴.

The Sm/Nd enrichment factor ($f^{\text{Sm/Nd}}$) of the Kambalda magma source can be calculated as a function of the time (T_s)

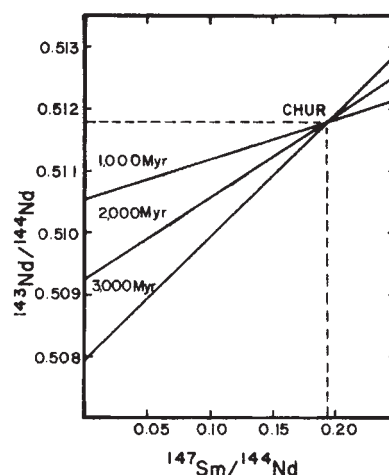


Fig. 4 Sm-Nd evolution diagram for isochrons of age 1,000, 2,000 and 3,000 Myr. Due to the approximate chondritic Sm/Nd evolution of the mantle^{9-12,18,24-26}, mantle-derived samples with Sm/Nd approximately the same as chondrites (CHUR) do not alone provide age information.

that it became fractionated relative to CHUR by using the relationship¹⁸:

$$\epsilon_{Nd}(T) = f^{Sm/Nd} Q_{Nd}(T_s - T)$$

For T_s equal to the age of the Earth ($T_s = 4,550$ Myr, $T = 2,790$ Myr, $Q_{Nd} = 0.0251$ Myr⁻¹), $f^{Sm/Nd} = 0.08$. This is a minimum value and implies a mantle source with a time average Sm/Nd from 4,550 Myr to 2,790 Myr that was 8% higher than CHUR. This evolution curve is shown in Fig. 5b together with the initial $^{143}Nd/^{144}Nd$ of Kambalda and the older mafic-ultramafic sequences, Talga-Talga³², Onverwacht¹¹ and Isua¹⁰. The colinearity of these data suggest the possible existence of an ancient (>4,000 Myr) ultramafic mantle source with $f^{Sm/Nd} \sim 0.08$. This reservoir would today have $\epsilon_{Nd} \sim +9$ and could thus be the source of the positive ϵ_{Nd} oceanic basalts. This is contrary to previous suggestions^{29,30} of a younger (1,800–2,000 Myr) source for modern day oceanic basalts. To distinguish between these evolutionary models it is apparent that additional high precision initial $^{143}Nd/^{144}Nd$ are required from Precambrian oceanic basalts of different ages.

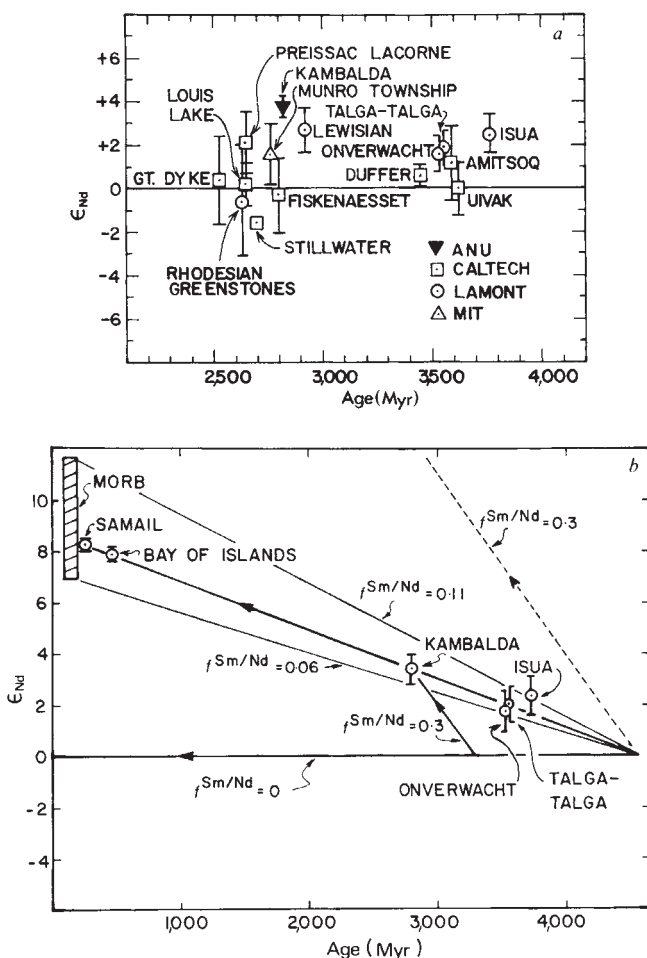


Fig. 5 *a*, Fractional deviations in parts per 10⁴ of initial $^{143}Nd/^{144}Nd$ of Archean rocks^{9-12,18-20,23-26,32}, from evolution in the CHUR reservoir¹⁸. The ϵ_{Nd} value of $+3.4 \pm 0.6$ for the Kambalda sequence implies derivation from a long-lived LREE depleted (high Sm/Nd) source. *b*, Possible evolution of $^{143}Nd/^{144}Nd$ based on the initial $^{143}Nd/^{144}Nd$ of Kambalda, Isua¹⁰, Onverwacht¹¹ and Talga-Talga³². The single evolution curves with $f^{Sm/Nd}$ of from 0.06 to 0.10 can account for the initial $^{143}Nd/^{144}Nd$ of these samples as well as present day oceanic basalts. An alternate explanation for the high initial $^{143}Nd/^{144}Nd$ of the Kambalda sequence involves multi-stage evolutions with $f^{Sm/Nd} \leq 0$. For comparison initial $^{143}Nd/^{144}Nd$ and inferred evolution of some highly fractionated ($f^{Sm/Nd} \approx 0.3$) Apollo 12 and 17 lunar maria basalt sources^{21,22} are also shown (dashed line).

An ancient (>4,000 Myr) high $f^{Sm/Nd}$ mantle source would imply an early differentiation of the Earth as well as incomplete mantle mixing to allow the preservation of these distinctive magma reservoirs. We note that this hypothesis is independent of the exact CHUR evolution curve for the bulk Earth but may be influenced by possible interlaboratory biases between the Kambalda and Onverwacht, Talga-Talga, Isua data. The initial $^{143}Nd/^{144}Nd$ of the Apollo 12 and 17 lunar maria basalts^{21,22} is also consistent with an early differentiation ($\sim 4,500$ Myr) of planetary bodies resulting in the formation of even more highly fractionated magma reservoirs having $f^{Sm/Nd} = 0.3$.

An alternative interpretation of the Kambalda initial ratio is that the time of differentiation of the source is substantially younger than 4,500 Myr. This would require a more strongly LREE depleted source. For example, for $T_s = 3250$ Myr, $f^{Sm/Nd} = 0.3$, which by analogy with the lunar data, is probably a maximum $f^{Sm/Nd}$ value.

Conclusions

The Sm-Nd whole rock systems have been examined in two volcanic sequences from the eastern Yilgarn Block of Western Australia. The predominantly mafic-ultramafic sequence at Kambalda has preserved uniform and regular Sm-Nd systems consistent with a well defined crystallization age of $2,790 \pm 30$ Myr and an initial $^{143}Nd/^{144}Nd = 0.50839 \pm 3$. In contrast, data from the felsic Kanowna complex are highly dispersed with a poorly defined age of $3,130 \pm 250$ Myr and initial $^{143}Nd/^{144}Nd = 0.50813 \pm 23$. The differences between the two sequences are poorly understood but may reflect disturbance of the Sm-Nd system or large variations in the age and initial ratios of the Kanowna complex. Due to the deviant Kanowna data array, the age difference between Kanowna and Kambalda is not considered to be significant although it is consistent with inferred stratigraphic correlations^{13,14}.

Although the whole rock samples from Kambalda form a perfect linear array (apart from an altered ultramafic) the age and initial $^{143}Nd/^{144}Nd$ is constrained by combining mafic-ultramafic volcanics with felsic volcanics. Consequently any age differences which may exist between felsic and mafic volcanics are not readily apparent. For example, a plausible hypothesis for the formation of the Kambalda complex may be the extrusion of the mafic-ultramafic volcanics in an ocean basin at $2,910 \pm 170$ Myr (age given by the mafic and ultramafic volcanics only). Then, within 120 ± 200 Myr (that is at 2,790 Myr) the sodic granite and related felsic intrusives were produced by partial melting of an eclogite source. An eclogite or more specifically a garnet bearing source²⁸ is implied by their pronounced LREE enrichments ($f^{Sm/Nd} \ll 0$) and may be the subducted equivalents of the mafic and ultramafic volcanics now preserved in the Kambalda complex.

Regardless of the question of the original age, the initial $^{143}Nd/^{144}Nd$ of the Kambalda sequence is significantly greater than the chondritic evolution curve with an ϵ_{Nd} value of $+3.4 \pm 0.6$. For the sodic-granite alone, the ϵ_{Nd} value is independently set at $+3.9$, if the Pb-Pb⁸ age of 2,760 Myr is used. The high initial $^{143}Nd/^{144}Nd$ requires an earlier LREE depleted (high Sm/Nd) source for the Kambalda volcanics and therefore provides some of the first definitive evidence for long-lived heterogeneities in the Archaean mantle.

We thank M. Perfit, O. A. Bavington and S. R. Taylor for REE analyses of the Kambalda ultramafics and hanging-wall basalt. J. J. Gresham, G. Loftus-Hills and J. C. Roddick provided the Kambalda samples and helped interpret the geology. We thank J. Barreda for typing the manuscript.

Received 1 July; accepted 6 October 1981.

1. Viljoen, R. P. & Viljoen, M. J. *Geol. Soc. S. Afr. Spec. Publ.* **2**, 275–296 (1969).
2. Green, D. H. *Earth planet. Sci. Lett.* **15**, 263–270 (1972).
3. Nesbitt, R. W. & Sun, S.-S. *Earth planet. Sci. Lett.* **31**, 433–453 (1976).
4. Sun, S.-S. & Nesbitt, R. W. *Earth planet. Sci. Lett.* **35**, 429–448 (1977).
5. Arriens, P. A. *Spec. Publ. J. geol. Soc. Aust.* **3**, 11–23 (1971).
6. Compston, W. & Turek, A. *J. geol. Soc. Aust.* **20**, 217–222 (1973).
7. Roddick, J. C. thesis, Australian Natn. Univ. (1974).
8. Oversby, V. M. *Geochim. cosmochim. Acta* **39**, 1107–1125 (1975).

9. Hamilton, P. J., O'Nions, R. K. & Evensen, N. M. *Earth planet. Sci. Lett.* **36**, 263–268 (1977).
10. Hamilton, P. J., O'Nions, R. K., Evensen, N. M., Bridgewater, D. & Allart, J. H. *Nature* **272**, 41–43 (1978).
11. Hamilton, P. J., Evensen, N. M., O'Nions, R. K., Smith, H. S. & Erlank, A. J. *Nature* **279**, 298–300 (1979).
12. Zindler, A., Brooks, C., Arndt, N. T. & Hart, S. *Int. Conf. Geochron. Cosmochron. Isotope Geol.* 469–471 (USGS Open File Rep. 78–701, 1978).
13. Williams, I. R. *Geol. Surv. West. Aust. A. Rep.* 40–41 (1969).
14. Gemuts, I. & Theron, A. in *Econ. Geol. of Aust. and Papua New Guinea*, Vol. 1. (ed. Knight, C. L.) 66–74 (The Australasian Institute of Mining and Metallurgy, Victoria, 1975).
15. Ross, J. R. & Hopkins, G. M. in *Econ. Geol. of Aust. and Papua New Guinea*, Vol. 1 (ed. Knight, C. L.) 100–121 (The Australasian Institute of Mining and Metallurgy, Victoria, 1975).
16. Barrett, F. M., Binns, R. A., Groves, D. I., Marston, R. J. & McQueen, K. G. *Econ. Geol.* **72**, 1195–1223 (1977).
17. McCulloch, M. T. & Chappell, B. W. *Earth planet. Sci. Lett.* (in the press) (1981).
18. DePaolo, D. J. & Wasserburg, G. J. *Geophys. Res. Lett.* **3**, 249–252 (1976).
19. McCulloch, M. T. & Wasserburg, G. J. *Science* **200**, 1003–1011 (1978).
20. DePaolo, D. J. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **43**, 999–1008 (1979).
21. Lugmair, G. W., Scheinin, N. B. & Marti, K. *Proc. 6th Lunar Sci. Conf.* 1419–1429 (1975).
22. Nyquist, L. E., Shih, C.-Y., Wooden, J. L., Bansal, B. M. & Wiesmann, H. *Proc. 10th Lunar Sci. Conf.* 77–114 (1979).
23. McCulloch, M. T. thesis, California Institute of Technology (1979).
24. DePaolo, D. J. & Wasserburg, G. J. *Geophys. Res. Lett.* **3**, 743–746 (1976).
25. Hamilton, D. J., Evensen, N. M., O'Nions, R. K. & Tarney, J. *Nature* **277**, 25–28 (1979).
26. Jacobsen, S. B. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **50**, 139–155, (1980).
27. Sun, S.-S. & Nesbitt, R. W. *Contr. Miner. Petrol.* **65**, 301–325 (1978).
28. Nance, W. B. & Taylor, S. R. *Geochim. cosmochim. Acta* **41**, 225–231 (1977).
29. Tatsumoto, M. *Science* **153**, 1094–1101 (1966).
30. Sun, S.-S. & Hanson, G. N. *Geology* **3**, 297–302 (1975).
31. Basu, A. R., Ray, S. L., Saha, A. K. & Sarkar, S. N. *Science* **212**, 1502–1506 (1981).
32. Hamilton, P. J., Evensen, N. M., O'Nions, R. K., Glikson, A. Y. & Hickman, A. H. *2nd Int. Archaeol. Symp. Perth Abstr.* 11–12 (1980).
33. Bavington, O. A. thesis, Australian Natn. Univ. (1979).
34. DePaolo, D. J. *Nature* **291**, 193–196 (1981).

Structure of vitamin D-dependent calcium-binding protein from bovine intestine

D. M. E. Szebenyi, S. K. Obendorf & K. Moffat

Section of Biochemistry, Molecular and Cell Biology, Clark Hall, Cornell University, Ithaca, New York 14853, USA

The structure of vitamin D-dependent calcium-binding protein from bovine intestine, determined crystallographically to 2.3 Å, contains a pair of calcium-binding domains similar to those in parvalbumin, but one domain has a longer and rearranged calcium-binding loop. This result has implications for structure predictions of other calcium-binding proteins such as calmodulin and S-100.

THE role of calcium as a 'second messenger' in the regulation of intracellular processes is now well established^{1,2}, and requires that intracellular calcium concentrations be precisely controlled. A wide variety of calcium-binding proteins, with affinity constants³ in the range 10^{-8} – 10^{-5} M, have been identified which either participate in this control or exhibit activities modulated by calcium (ref. 4 and refs therein). One such class of proteins comprises the vitamin D-dependent calcium-binding proteins, which have been studied in various species. In birds, a protein of molecular weight (MW) of ~28,000 is found in the intestine, kidney, brain, shell gland and other organs. In mammals, a protein of similar size is found in various tissues, but the predominant intestinal calcium-binding protein is a smaller, 10,000-MW species⁵. The chicken protein has four high-affinity calcium-binding sites ($K_D \sim 10^{-6}$ M) and the bovine intestinal protein two such sites^{5,6}. The intestinal calcium-binding proteins are soluble species, located in the cytoplasm of the absorptive cells⁷. Their synthesis requires the presence of 1α , 25-dihydroxyvitamin D₃ ($1,25(\text{OH})_2\text{D}_3$) which, when given to a rachitic rat, produces an immediate increase in intestinal absorption of calcium, followed after a delay of several hours by the production of calcium-binding protein⁸. The protein thus cannot be involved in the initial translocation of calcium across the gut wall but could play a part in later stages of calcium metabolism. Alternatively, it could act as an intracellular calcium buffer, as the calcium-sensitive component of a transport molecule or as some other calcium-controlled regulator of absorptive cell function⁹.

Although the three-dimensional structures of several calcium-binding proteins have been determined, of these only carp parvalbumin is believed to be involved in the regulation of intracellular calcium³. Parvalbumin contains two 29-amino acid calcium-binding domains, each consisting of a loop of 12 residues flanked by two α -helices approximately at right angles to each other^{10,11}. This helix-loop-helix conformation is termed¹¹ an 'EF hand' (see Fig. 1 of ref. 12 for derivation of the name). Based on an analysis of the key structural features of the two EF hands in parvalbumin, Tufty and Kretsinger¹² identified 16 amino acids in the 29-residue sequence that they believed were essential to the formation of a calcium-binding EF hand. They applied this 16-residue test sequence to a wide range of protein

sequences, and predicted that EF hands would be found in troponin C and other muscle proteins¹². Subsequently this prediction was generalized to: intracellular calcium-modulated proteins contain EF hands and conversely, any protein which contains an EF hand is calcium modulated^{3,13,14}.

This hypothesis has attracted broad interest but has not been subjected to a structural test. We present here the three-dimensional structure of a vitamin D-dependent calcium-binding protein from bovine intestine (ICaBP), which from its amino acid sequence⁶ is predicted to contain one, and possibly two, EF hand(s) (see Fig. 1). This protein indeed contains two helix-loop-helix calcium-binding domains, resembling the EF hands in parvalbumin in both internal conformation and relationship to one another, but differing from them in detail.

X-ray analysis

ICaBP (minor A form; see Fig. 1) was purified¹⁵ from bovine intestinal mucosa to electrophoretic homogeneity and crystallized at pH 8.7–8.9 (ref. 16, Table 1). Derivatives for multiple isomorphous replacement phasing were prepared by soaking crystals in solutions containing various heavy-atom reagents and X-ray data were collected using both screened precession films and a Syntex P2₁ diffractometer, as summarized in Table 1. The single site in the Nd³⁺ derivative was located by a difference Patterson synthesis; sites in subsequent derivatives were determined by a combination of difference Patterson and cross-phased difference Fourier syntheses¹⁷. Heavy-atom refinement was carried out using a program based on the ideas of Blow and Crick¹⁸, and of Matthews¹⁹. Statistics for the final cycle of refinement are shown in Table 1.

The native electron density map calculated at 2.3-Å resolution using the multiple isomorphous replacement phases shows most of the molecular boundary and the general course of the polypeptide chain but appears noisy. To increase the signal-to-noise ratio, we applied a density modification procedure based on those of Schevitz *et al.*²⁰ and of Agard and Stroud²¹. This approach imposes two constraints: non-negativity of the electron density and uniformity (at a low level) of the electron density in solvent regions. Although only 35% of the unit cell is solvent, compared with 82% in the tRNA crystals used elsewhere²⁰, this proved a useful approach. Density modification

	Helix	Loop	Helix	Linker
Test sequence	E L _ _ L L _ _ L	* _ * _ * G _ I _ *	L _ _ L L _ _ L	
Parvalbumin (carp muscle, component 3)	...38K S A D D V K K A F A I I	D Q D K S G F I E E D E	L K L F L Q N F	K A D A R A L T D G ₈₀
	81E T K T F L K A G	D S D G D G K I G V D E	F T A L V K A ₁₀₈	
ICaBP (bovine intestine, minor A)	1K S P E E E L K G I F E K Y A	K E G D F Q L S K E E	L K L L L Q T E	F P S L L K G P S ₄₄
	45T L D E L F E E L D	K N G D G E V S F E E	F Q V L V K K I	S ₇₅
S-100 PAPI-b (bovine brain)	1E S E L K A V V A L I D V F H Q Y	G R E G D K K L K K S E	L K E L I N N E	L S H F L E E I K E Q E ₅₁
	52V V D K V M E T L D	S D G D G E C D F Q E	F M A F V A M I	T T A C H E F F E H E ₉₁
Troponin C (rabbit skeletal)	...14E M I A E F K A A F D M F	D A D G G G D I S V K E	L G T V M R M L	G Q T P T R E ₅₃
	54E L D A I I E E V D E D G S G T I D F E E	F L V M M V R Q	M K E D A K G K S E	E ₉₃
	94E L A E C F R I F D R N A D G Y I D A E E	L A E I F R A S	G E H V T D E ₁₂₉	
	130E I E S L M K D G D K N N D G R I D F D E	F L K M M E G V	Q ₁₅₉	
Calmodulin (bovine brain)	...7E Q I A E F K E A F S L F	D K D G N G T I T T K E	L G T V M R S L	G Q N P T E A ₄₆
	47E L Q D M I N E V D A D G N G T I D F P E	F L T M M A R K	M K D T D S E E ₈₃	
	84E I R E A F R V F D K D G N G Y I S A A E	L R H V M T N L	G E K' L T D E ₁₁₉	
	120E V D E M I R E A N I D G D G E V N Y E E	F V Q M M T A K ₁₄₈		

Fig. 1 Sequences of some calcium-binding proteins, with EF hand test sequence. The test sequence represents residues suggested from earlier studies to be critical to the EF hand structure¹². In the test sequence, * indicates an oxygen-containing residue (D, E, N, Q, S, T), L indicates a hydrophobic residue (L, V, I, F, M), I indicates a hydrophobic residue with isoleucine preferred, G is glycine, and an underscore indicates that any amino acid may appear in that position. Insertions and deletions have been postulated to maximize homology between sequences. N-terminal portions of parvalbumin³⁴, troponin C³⁵ and calmodulin³⁶ have been omitted. The 'major' (native) form of ICaBP has the sequences Ac-Ser-Ala-Lys-minor A. Amino acid code: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; K', trimethyllysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine. S-100 PAPI-b (bovine brain) sequence was from ref. 37; that of ICaBP, ref. 6.

Table 1 Statistics for multiple isomorphous replacement phasing of ICaBP

	Native	NdCl ₃ (1 mM; 2-day soak)	KAu(CN) ₂ (5 mM; 3-14-day soak)	K ₂ Pt(CN) ₄ (5 mM; 3-day soak)
Unique reflections	3,023	3,031	3,124	2,601
Scaling <i>R</i>	0.0687	0.0795	0.0754	—
Film data resolution (Å)	2.3	2.3	2.3	—
Diffractometer data resolution (Å)	2.9	—	3.4	2.7
Reflections used	2,157*	2,025	2,051	1,532
No. of sites		1	2	10
Occupancy (electrons)		19†	45, 43	10, 12, 16, 15, 13 11, 8, 12, 7, 7
<i>R_c</i>		0.3441	0.5770	0.4098
<i>R_K</i>		0.0718	0.1444	0.1553
<i>E/f</i>		0.7909	0.6018	1.0826
A.D. residual		2.00	1.53	12.33
<i>m</i> , final refinement cycle	0.6446			
<i>m</i> , phasing of native map	0.6109			
<i>m</i> , after density modification	0.7872			

$$R_c = \frac{\sum |F_{pH_{obs}}| - |F_{pH_{calc}}|}{\sum |F_p| - |F_{pH_{obs}}|}, \text{ summed over centric reflections}$$

$$R_K = \frac{\sum |F_{pH_{obs}}| - |F_{pH_{calc}}|}{\sum |F_{pH_{obs}}|}, \text{ summed over all reflections}$$

Native crystals were grown from 65–80% saturated (NH₄)₂SO₄ in 0.02 M Tris pH 8.7–8.9, and stored in 90% saturated (NH₄)₂SO₄. Heavy-atom solutions contained the appropriate reagent in 90% saturated (NH₄)₂SO₄. The crystals are space group P2₁2₁2₁; unit cell dimensions are: *a* = 56.2 ± 0.1, *b* = 42.6 ± 0.1, *c* = 29.2 ± 0.1 Å. There is one molecule per asymmetric unit and a solvent content of 35%. Films were scanned on an Optronics P200 rotating drum densitometer and scaled together using the CORRELATE program of G. N. Reeke's ROCKS computing package. Diffractometer data were scaled to the film data; to compensate for absorption effects, different scale factors were permitted for different ranges of the diffractometer setting angles ϕ and χ . Intensities were placed approximately on an absolute scale using a Wilson plot. Derivative scale factors were adjusted to eliminate peaks and holes at heavy-atom sites on a native electron density map. Anomalous dispersion data for all derivatives were used in phasing. Nd³⁺ and Au(CN)₂⁻ positions and occupancies were refined together; Pt(CN)₄²⁻ positions and occupancies were refined phasing on all three derivatives but holding Nd³⁺ and Au(CN)₂⁻ parameters fixed. Effective isotropic temperature factors (*B*s) were set to values of 10–15 and not refined.

$$\text{A.D. residual} = \left[\frac{\sum w(AD_{obs} - AD_{calc})^2}{\sum wAD_{calc}} \right]^{1/2}$$

* Very weak native reflections and those with less than two good derivative measurements (counting the two members of a KAu(CN)₂ anomalous pair as two derivatives) were not used in phasing. Only the 2,080 reflections between 9.09 and 2.33 Å resolution were used for heavy-atom refinement, but all 2,157 reflections were used to phase the native map.

† Net occupancy. Nd³⁺ replaces Ca²⁺ (18 electrons), so the total Nd³⁺ occupancy would be ~37 electrons.

increased the figure of merit from 0.61 to 0.79 and improved the clarity of the electron density map (compare the representative electron density sections before modification [Fig. 2a] with those after [Fig. 2b]). One or two spurious breaks in the polypeptide chain were introduced, but in most regions the modified map proved easier to interpret.

A Kendrew-Watson molecular model of ICaBP was constructed from the modified map using a Richards optical comparator²². The polypeptide backbone is shown in Fig. 3, based on the α -carbon positions measured from this model. In most places, the course of the main chain is clear, although the exact conformation in some of the non-helical regions remains to be confirmed by crystallographic refinement. The register of the amino acid sequence⁶ with the map is firmly established by the placement of the four prolines, all of which appear as flattened regions of high electron density in the main chain, and

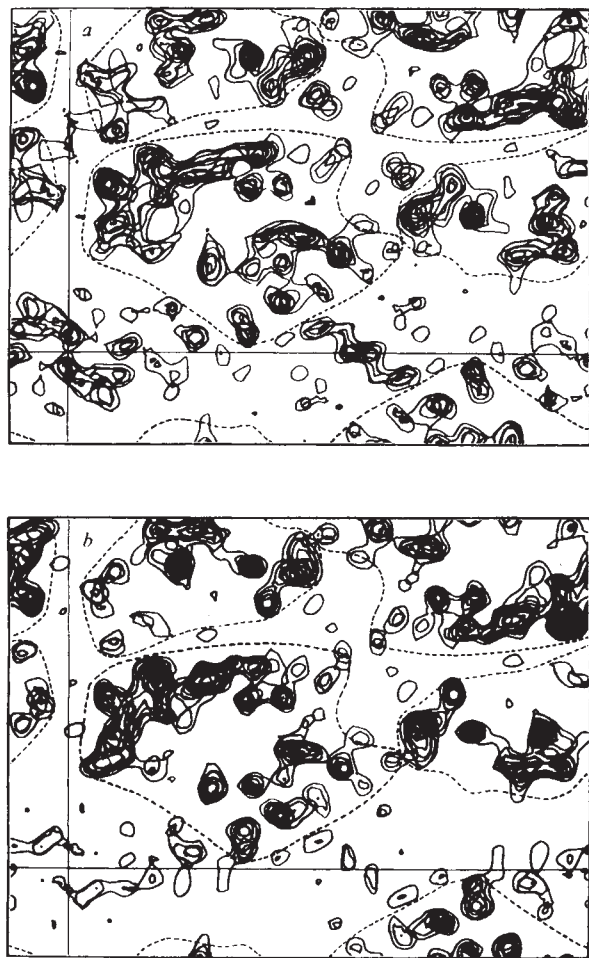


Fig. 2 Section through ICaBP electron density, from $z = 0.375$ to $z = 0.425$, before (a) and after (b) density modification. The density modification procedure was as follows: a probable molecular envelope was determined by inspection of the initial map, and all electron density values outside the envelope were considered to be in solvent regions and were set to a small constant value. Within the envelope, negative values were multiplied by 0.1 (to flatten out holes without creating sharp edges) and peaks above a certain maximum (slightly higher than the largest value on the original map) were truncated to that maximum. To reduce artefacts due to sharp edges, each point just outside the molecular envelope or just within the rim of a truncated peak was set to the average of itself with the 26 surrounding points. Inverse Fourier transformation of the modified map produced a new set of structure factors, whose phases were combined with the original phases using the method of Brice³⁸. The combined phases were used with the observed amplitudes to produce a new map. After four cycles, the envelope was adjusted as indicated by the resulting map and applied to the original map. Seven cycles were then carried out using the revised envelope (indicated by dotted lines).

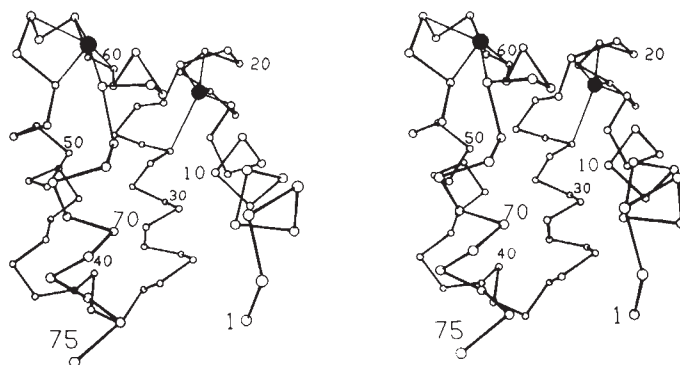


Fig. 3 Stereo view of the ICaBP α -carbon backbone, with the two bound calcium ions. $\circ$, α -carbons; $\bullet$, calcium ions.

by very clearly defined density for all aromatic side chains. The density is not consistent with an earlier sequence²³.

Overall structure of ICaBP

ICaBP is a globular protein of approximate dimensions $25 \times 30 \times 30$ Å. It consists of four helices, designated I–IV and comprising residues 2–14, 24–35, 46–54 and 63–75, respectively, wrapped around the usual hydrophobic core of a soluble protein. There are hydrophobic contacts between side chains from all pairs of helices, and between the II–III linker and helices II and III. In addition to the four main helices, residues 36–41 are approximately helical, with hydrogen bonds formed between 36–40 and 38–41. Even neglecting this region, the helix content is $\sim 63\%$, higher than indicated by circular dichroism studies ($40 \pm 5\%$ for bovine ICaBP (M. Jibson, unpublished results); 30% for the closely homologous porcine protein²⁴). Regions I, II and III are standard α -helices whereas IV begins with four residues of 3_{10} helix and then becomes α -helical. The region between helices III and IV is a calcium-binding loop of the EF hand type. Between I and II is a slightly longer loop which also binds calcium and may be described as a variant EF hand (see below). If the linker region between II and III is neglected, there is an approximate 2-fold axis relating the I–II and III–IV domains, illustrated by the view down this axis shown in Fig. 4a.

The single tyrosine, Tyr 13, lies on the surface of the molecule, making a probable hydrogen bond to Glu 35, thus tying helices I and II together. Birdsall *et al.*²⁵ proposed, on the basis of NMR work, that the Tyr lies in a hydrophobic pocket involving one Ile and two Phe. We find Tyr 13 to be in a very shallow depression formed by Ile 9, Leu 30, Leu 31, Thr 34 and Glu 35. Phe 10, Phe 36 and Phe 66 are within 11 Å of the Tyr, which is close enough for some interaction. Fluorescence measurements demonstrate complete resonance energy transfer from all five Phe residues to the Tyr²⁶. For the two Phe further than 10–12 Å from the Tyr, energy transfer presumably involves an inter-Phe relay system; no Phe is more than 11 Å from the rest.

EF hands in ICaBP

The calcium-binding loop between helices III and IV is extremely similar to the EF hand prototype found in parvalbumin. The calcium is coordinated by side-chain oxygens from Asp 54, Asn 56, Asp 58 and Glu 65 and by the main-chain carbonyl from Glu 60. Both carboxyl oxygens of Glu 65 are ligands, so that there are six protein oxygens surrounding the calcium ion at distances (unrefined) of 2.2–2.6 Å. The calcium in the EF loop of parvalbumin has a water molecule in its coordination sphere; the ICaBP III–IV calcium is clearly exposed to solvent, and initial refinement suggests that a solvent molecule is a seventh ligand. Helices III and IV are at an angle of 121° with a minimum inter-axis distance of 8.06 Å, compared with 102° and 7.63 Å in the parvalbumin EF hand. Fitting the α -carbons of the ICaBP III–IV domain to those of the parvalbumin EF region

Table 2 ICaBP compared with parvalbumin

<i>a</i> Inter-helix angles and distances					
Parvalbumin			ICaBP		
Helix pair	Inter-helix angle	Minimum distance (Å) between helix axes	Helix pair	Inter-helix angle	Minimum distance (Å) between helix axes
EF hand helices					
C-D	103°	8.37	I-II	121°	8.06
E-F	102°	7.63	III-IV	121°	8.55
Other helices in contact					
A-B	155°	5.06			
A-D	150°	10.35			
A-E	120°	9.34			
B-F	73°	7.80			
C-F	119°	9.35	I-IV	123°	10.87
D-E	117°	9.59	II-III	121°	9.90
			II-IV	21°	11.52
<i>b</i> EF hand α -carbon fits					
Segments compared			No. of atoms	r.m.s. discrepancy (Å)	r.m.s. discrepancy 14 loop atoms (Å)
CD (40-69) vs. EF (79-108)			30	1.86*	0.93
CD (40-71) vs. I-II (3-36) [†]			32	2.86	1.99
EF (78-108) vs. I-II (2-34) [†]			31	2.75	2.11
CD (43-71) vs. III-IV (46-74)			29	3.25	1.41
EF (82-108) vs. III-IV (46-72)			27	2.40	1.26
I-II (6-36) [†] vs. III-IV (46-74)			29	3.33*	2.43
CD, EF (38-71, 82-108) vs. I-II, III-IV (1-36 [†] , 46-72)			61	3.07	

a, Angles and minimum inter-axial distances between all pairs of helices which are in contact. The inter-helix angle is calculated by considering the helix axes as vectors, and is 0–90° for parallel, 90–180° for antiparallel helices. The distance is the minimum distance between infinite lines along the helix axes; the closest approach of the finite helices may be greater. b, Agreement between α -carbon positions of indicated segments after rotation and translation of one set of coordinates so as to minimize $\sum d_i^2$, where d_i is the distance between the i th α -carbons of the two segments after rotation and translation. Parvalbumin coordinates were supplied by the Protein Data Bank³³.

* The transformation relating these segments, expressed as a screw operation, is 172°, –0.78 Å for parvalbumin, and 168°, –0.32 Å for ICaBP, that is, in both cases the two segments are related by an approximate 2-fold axis.

† Asn 21 omitted. Ala 14 used in 14 residue loop but omitted from helix-loop-helix comparison.

results in an average (r.m.s.) discrepancy of 2.40 Å for the entire domain and 1.26 Å for the calcium-binding loops alone; the corresponding values for CD as opposed to EF are 1.86 and 0.93 Å (Table 2).

However, the I-II loop (two residues longer) is rather different. The insertion of Ala 14 means that the first predicted calcium ligand, Ala 15, is on the outside rather than the inside of helix I, and of course the alanine side chain contains no oxygen. This lack is accommodated by using the main-chain carbonyl of Ala 15 as a ligand. The loop is further rearranged about the second and third calcium ligands, Glu 17 and Asp 19, so that they also coordinate Ca^{2+} with their main-chain carbonyls rather than their side chains. Following the second insertion of Asn 21 (or Pro 20), the loop adopts essentially the normal EF hand conformation. The main-chain carbonyl of Glu 22 and the two carboxyl oxygens of Glu 27 constitute the remaining calcium ligands, giving a total of six protein oxygen ligands. Ca–O distances (unrefined) are 2.2–2.6 Å. Due to the different conformation of the N-terminal portion of the loop, the I-II calcium

is much less accessible to solvent than the III-IV calcium, and it is unlikely that a solvent molecule is a seventh ligand.

The stretches of calcium-binding loop from residues 22–25 and 60–63 run antiparallel to one another and are joined by a hydrogen bond between Leu 23 and Val 61, across the approximate 2-fold axis, resembling a small piece of β -sheet. This arrangement also occurs in parvalbumin, between the CD and EF loops. The calcium ions are separated by 11.9 Å in parvalbumin and by 12.5 Å in ICaBP.

In both the conformation of the two calcium-binding domains and their disposition with respect to each other, ICaBP thus resembles the C-terminal two-thirds of parvalbumin (Fig. 4). If the II-III linker and the two inserted residues in the I-II loop are omitted, a fit of the remaining 61 ICaBP α -carbons to the corresponding parvalbumin α -carbons gives a r.m.s. discrepancy of 3.07 Å. Compare, for example, the 5.7 Å r.m.s. discrepancy between 60-residue segments of T4 lysozyme and hen egg-white lysozyme, and the 0.9 Å (for a 40-residue stretch) to 1.5 Å (for 80 residues) discrepancy between myoglobin and haemoglobin β -chain²⁷.

Despite this resemblance, there are significant differences between the conformations of ICaBP and parvalbumin. Helices II and IV in ICaBP are in contact, whereas the corresponding helices D and F in parvalbumin are not (Table 2). Inter-helix angles differ and are generally more regular in ICaBP (Table 2). Parvalbumin has an extended D–E linker, a bend in the D helix and a straight F helix, whereas the ICaBP II–III linker is highly folded, helix II is straight and there is a bend in helix IV. These differences in helix spacing and linker folding are only to be expected, because if ICaBP were to adopt the parvalbumin conformation, it would have a large hole in its 'underside' which is filled by the A and B helices in parvalbumin. ICaBP helix III is shorter than parvalbumin helix E and IV is slightly longer than F. Finally, the I-II and CD loops differ: the latter is a normal EF hand calcium-binding loop whereas the former is a variant. Thus, although ICaBP exhibits the same overall helix-loop-helix structure found in parvalbumin, the disposition of the

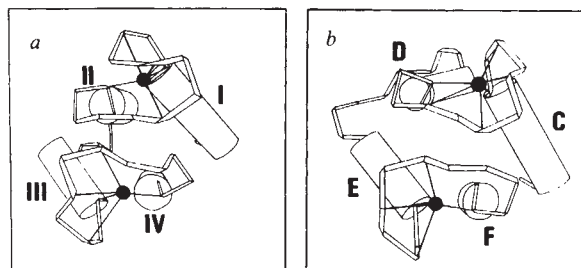


Fig. 4 α -Carbon backbone and calcium ions viewed down the approximate intramolecular 2-fold axis of (a) ICaBP and (b) parvalbumin, residues 38–108. Helices are represented by cylinders and other polypeptide chains as a folded ribbon. ●, Calcium ions.

helices with respect to each other, both within one domain and between domains, is clearly different. This may hinder attempts to search for EF hands by application of molecular replacement techniques to crystals of other calcium-regulated proteins such as calmodulin and troponin C.

Calcium-binding to ICaBP

We believe that the structure described here represents ICaBP having both high-affinity sites fully loaded with calcium and no significant occupancy of low-affinity calcium sites. Weakly bound calcium was removed by dialysis against distilled water. The presence of calcium in the high-affinity sites is evidenced by peaks near both sites on a native anomalous difference Fourier (the two highest peaks on the map); and, in I-II, by the presence of a peak at the calcium position on the native electron density map. Site III-IV does not have a peak at exactly the calcium position, but as it is a heavy-atom site (for the Nd^{3+} derivative), distortions in this region are to be expected. Full occupancy of the high-affinity sites is suggested by the good definition of the polypeptide chain in both calcium-binding loops and by the absence of changes in structure amplitudes on soaking crystals in 10 mM CaCl_2 . NMR²⁵ and trypsin digestion²⁸ experiments reveal a substantial conformational change in the protein on removal of one or both calcium ions. This conformational change is unfortunately sufficient to disintegrate ICaBP crystals soaked in 1.0 mM EDTA, and attempts to grow crystals in the absence of calcium have been unsuccessful. Thus, the conformational changes which occur on release of one or two calcium ions are unknown; they might involve only small rearrangements of the calcium-binding loops, or motion of the flanking helices with respect to each other, or motion of the two calcium-binding domains.

The two calcium-binding sites differ in their calcium affinities, as indicated by the removal in EDTA (solution of 1:1 EDTA/ICaBP) of only one calcium (C. S. Fullmer, personal communication). The relative affinities for Tb^{3+} , a fluorescent model for Ca^{2+} , also differ, by at least a factor of 20 (ref. 26). It is apparent from the X-ray structure that the III-IV site is more exposed to solvent than the I-II site; this agrees with the observation that, in a crystal soaked for 2 days in 1.0 mM NdCl_3 , Nd^{3+} displaces the III-IV but not the I-II Ca^{2+} . We suggest that the I-II site may be a structural site, saturated with calcium in all physiological conditions, whereas the III-IV site is a regulatory site which binds or releases calcium as intracellular calcium levels vary.

Implications for the structure of other calcium-binding proteins

ICaBP is a member of the calmodulin family of calcium-binding proteins, for which it is proposed that an ancestral four-domain calcium-binding protein was produced by successive gene duplications²⁹. Calmodulin and troponin C have retained this type of structure, and bind four calcium ions per protein molecule. Some other members of the family have lost the calcium-binding ability of one or more of the domains. In parvalbumin, the first domain was deleted entirely and the second, the AB domain, no longer binds calcium; in ICaBP, two domains were deleted and the N-terminal I-II domain was modified by the insertion of two residues, while retaining its calcium-binding ability.

The S-100 PAPI-a and PAPI-b proteins belong to the same branch of the calmodulin family as ICaBP; divergence seems to have occurred after the modification of the I-II loop (Fig. 1 shows the PAPI-b sequence; PAPI-a is 58% homologous to PAPI-b overall and almost identical with respect to the calcium-binding loops³⁰). The S-100 proteins are thus expected to contain one variant and one normal EF hand, arranged similarly to those in ICaBP.

The calcium-binding characteristics of PAPI-b, however, are more complex than those of ICaBP. Two Ca^{2+} are bound (per monomer) at pH 8.3 but only one at pH 7.6. Furthermore, the

presence of K^+ inhibits binding of calcium to the high-affinity sites³¹. PAPI-b contains three basic side chains Lys-His-Lys in positions corresponding to Pro-Asn-Gln in the I-II loop of ICaBP (Fig. 1). It is likely that titration of the histidine is responsible for the pH dependence of calcium-binding, and the overall affinity of this site for cations may be lowered relative to ICaBP by the proximity of additional basic side chains.

Kretsinger and Barry³² constructed a model for the four-domain calcium-binding protein, troponin C, from four EF hands, by first pairing two hands as in the parvalbumin CD plus EF region, duplicating the pair and placing the non-calcium-binding ends of the pairs together so that a quasi 2-fold axis perpendicular to the one within each pair relates the two pairs. The molecule then exhibits quasi-222 symmetry. A similar model may be constructed for other four-domain proteins such as calmodulin. This structural prediction is supported by our results: the conformation of the ICaBP III-IV domain proves that the normal EF hand occurs other than in parvalbumin, and the stability of pairs of EF hands is demonstrated by the maintenance of such a pair, related by an approximate 2-fold axis, in ICaBP even when one calcium-binding loop has been appreciably modified.

The overall conformations of proteins in the calmodulin family are thus determined by the preferred formation of EF hands and pairs of hands, which may be packed together into higher aggregates. Differences in function are made possible by variation in the conformation of the calcium-binding loops, in the disposition of helices with respect to each other, and/or by differences in the linker regions. The conformation of the linkers, in particular, is quite free to evolve as required for interaction with other molecules, without disturbing the overall shape or calcium-binding abilities of the protein.

This determination of the ICaBP structure provides the first structural confirmation of the EF hand hypothesis, and also expands the definition of an EF hand, first to include structures whose helices are disposed rather differently from those in parvalbumin (for example, the ICaBP III-IV domain), and second to include structures with a modified calcium-binding loop such as the ICaBP I-II domain. These latter variant EF hands are not necessarily detectable from sequence comparisons; if the existence of two insertions is not recognized, only 8 out of 16 residues of the I-II domain match the EF hand test sequence. Hence the recognition of such variants in other proteins (except for those such as the S-100 proteins which are clearly closely related to ICaBP) will require the determination of their three-dimensional structures.

We thank R. H. Wasserman and C. S. Fullmer for providing us with our initial samples of ICaBP, for assistance in protein preparation, and for the ICaBP sequence before its publication; J. Clardy made diffractometer time available, and B. W. Siegel contributed the use of his Optronics film densitometer. We received computer programs from various sources, notably G. N. Reeke, L. Ten Eyck, R. C. Ladner, G. Petsko and the CRYNET facility at Brookhaven. A. J. S. Jones improved the ICaBP purification scheme and performed the solution studies of ICaBP fluorescence and $\text{Tb}^{3+}/\text{Ca}^{2+}$ binding. J. Wenban provided assistance in the preparation of figures and in much of the experimental work. This research was supported by NIH grant AM 21453. K. M. holds a NIH Career Development Award, AM 00322.

Received 10 June; accepted 27 September 1981.

1. Rasmussen, H., Goodman, D. B. P. & Tenenhouse, A. *CRC Crit. Rev. Biochem.* **1**, 95-148 (1972).
2. Rasmussen, H., Jensen, P., Lake, W., Friedmann, N. & Goodman, D. B. P. *Adv. Cyclic Nucleotide Res.* **5**, 375-394 (1974).
3. Kretsinger, R. H. *A. Rev. Biochem.* **45**, 239-266 (1976).
4. Cheung, W. Y. *Science* **207**, 19-27 (1980).
5. Wasserman, R. H., Fullmer, C. S. & Taylor, A. N. in *Vitamin D* (ed. Lawson, D. E. M.) 133-136 (Academic, New York, 1978).
6. Fullmer, C. S. & Wasserman, R. H. *J. biol. Chem.* **256**, 5669-5674 (1981).
7. Taylor, A. N. in *Calcium Binding Proteins: Structure and Function* (eds Siegel, F. L. et al.) 393-400 (Elsevier, New York, 1980).
8. Thomasset, M., Cuisinier-Gleizes, P. & Mathieu, H. *FEBS Lett.* **107**, 91-94 (1979).

9. Wasserman, R. H. in *Calcium Binding Proteins: Structure and Function* (eds Siegel, F. L. et al.) 357–361 (Elsevier, New York, 1980).
10. Kretsinger, R. H. & Nockolds, C. E. *J. biol. Chem.* **248**, 3313–3326 (1973).
11. Moews, P. C. & Kretsinger, R. H. *J. molec. Biol.* **91**, 201–228 (1975).
12. Tufty, R. M. & Kretsinger, R. H. *Science* **187**, 167–169 (1975).
13. Kretsinger, R. H. in *Calcium Binding Proteins and Calcium Function* (eds Wasserman, R. H. et al.) 63–72 (Elsevier, New York, 1977).
14. Kretsinger, R. H. *CRC Crit. Rev. Biochem. Acta* **317**, 172–186 (1973).
15. Fullmer, C. S. & Wasserman, R. H. *Biochim. biophys. Acta* **317**, 172–186 (1973).
16. Moffat, K., Fullmer, C. S. & Wasserman, R. H. *J. molec. Biol.* **97**, 661–664 (1975).
17. Blundell, T. L. & Johnson, L. N. *Protein Crystallography* (Academic, London, 1976).
18. Blow, D. M. & Crick, F. H. C. *Acta crystallogr.* **12**, 794–802 (1959).
19. Matthews, B. W. *Acta crystallogr.* **20**, 82–86 (1966).
20. Schevitz, R. W. et al. *Nature* **278**, 188–190 (1979).
21. Agard, D. A. & Stroud, R. M. *Acta crystallogr.* (in the press).
22. Richards, F. M. *J. molec. Biol.* **37**, 225–230 (1968).
23. Huang, W.-Y., Cohn, D. V., Hamilton, J. W., Fullmer, C. S. & Wasserman, R. H. *J. biol. Chem.* **250**, 7647–7655 (1975).
24. Dorrington, K. J., Hin, A., Hofmann, T., Hitchman, A. J. W. & Harrison, J. E. *J. biol. Chem.* **249**, 199–204 (1974).
25. Birdsall, W. J. et al. in *Calcium Binding Proteins: Structure and Function* (eds Siegel, F. L. et al.) 405–406 (Elsevier, New York, 1980).
26. Jones, A. J. S., Szebenyi, D. M. & Moffat, K. in *Calcium Binding Proteins: Structure and Function* (eds Siegel, F. L. et al.) 413–416 (Elsevier, New York, 1980).
27. Remington, S. J. & Matthews, B. W. *J. molec. Biol.* **140**, 77–99 (1980).
28. Fullmer, C. S. & Wasserman, R. H. in *Calcium Binding Proteins and Calcium Function* (eds Wasserman, R. H. et al.) 303–312 (Elsevier, New York, 1977).
29. Goodman, M. in *Calcium Binding Proteins: Structure and Function* (eds Siegel, F. L. et al.) 347–354 (Elsevier, New York, 1980).
30. Isobe, T. & Okuyama, T. *Eur. J. Biochem.* **116**, 79–86 (1981).
31. Calissano, P., Alemá, S. & Fasella, P. *Biochemistry* **13**, 4553–4560 (1974).
32. Kretsinger, R. H. & Barry, C. D. *Biochim. biophys. Acta* **405**, 40–52 (1975).
33. Bernstein, F. C. et al. *J. molec. Biol.* **112**, 535–542 (1977).
34. Nockolds, C. E., Kretsinger, R. H., Coffee, C. J. & Bradshaw, R. A. *Proc. natn. Acad. Sci. U.S.A.* **69**, 581–584 (1972).
35. Collins, J. H., Potter, J. D., Horn, M. J., Wilshire, G. & Jackman, N. *FEBS Lett.* **36**, 268–272 (1973).
36. Watterson, D. M., Sharief, F. S. & Vanaman, T. C. *J. biol. Chem.* **255**, 962–971 (1980).
37. Isobe, T. & Okuyama, T. *Eur. J. Biochem.* **89**, 379–388 (1978).
38. Bricogne, G. *Acta crystallogr.* **A32**, 832–847 (1976).

LETTERS TO NATURE

Absorption of γ rays in active galaxies as a test of the jet hypothesis

L. Bassani & A. J. Dean

Department of Physics, University of Southampton,
Southampton SO9 5NH, UK

Powerful X-ray emission is a common feature of active galaxies, some of which have also been detected as γ -ray sources^{1–5}. The observed high luminosities, coupled with reports of variability on time scales of days or less at X-ray energies⁶ and of 1 yr or less at γ -ray energies³, suggest that both emissions may originate in the same volume deep within the nucleus of the galaxy. Under this assumption we investigate here the attenuation of γ rays as a result of pair-production in photon-photon collisions with X rays and show that whilst Seyfert galaxies are transparent to γ radiation up to ~ 1 GeV, QSOs and BL Lac objects are generally opaque about 1 MeV. However, the detection of γ rays from these latter classes of objects may be explained if beaming of the X- γ emission is assumed.

The relevance of photon-photon absorption to specific astronomical sources has been discussed by several authors^{7–9}. Following their formalism, a photon of energy E_X may collide with a photon of energy E_γ and produce an electron-positron pair provided that

$$E_X E_\gamma \geq \frac{2(mc^2)^2}{(1 - \cos \theta)} \quad (1)$$

where $mc^2 = 0.511$ MeV is the rest mass energy of the electron and θ the angle between the directions of motion of the two photons. For fixed values of E_γ , the pair production cross-section¹⁰ $\sigma(E_X)$ rises steeply from zero at the threshold energy $E_X = E_s$ to a maximum value $\sigma_{\max} = 1.7 \times 10^{-25}$ cm² at $E_X = 2E_s$ and then falls off as E_X^{-2} for energies $E_X \gg E_s$. The optical depth for this process can be written as:

$$\tau = R \int \sigma(E_X) N_X(E_X) dE_X \quad (2)$$

where the cross-section σ and the density of X-ray photons per unit energy N_X , are assumed to be constant throughout the source region of size R . In the first instance we consider the case of isotropic emission and for simplicity set $\theta = \pi/2$. Furthermore within the limits of accuracy required by this calculation it is sufficient to approximate $\sigma(E_X)$ by a rectangular function of height $\sigma_{\max}$ and width $2.5 E_s$. Consequently, the optical depth may be written as

$$\tau = R \sigma_{\max} N_X(2E_s) 2.5 E_s \quad (3)$$

To estimate N_X we use the X-ray luminosity and write

$$N_X(2E_s) = f(\alpha) L_X \frac{(2E_s)^{-\alpha}}{4\pi R^2 c} \quad (4)$$

where α is the source photon spectral index at X-ray energies.

For L_X taken over a fixed energy band (generally 2–10 keV), $f(\alpha)$ is a slowly varying function of α . Introduction of numerical values in equation (3) leads to the final expression

$$\tau = 2.5 \times 10^{-28} f(\alpha) \frac{L_X}{R} (2E_s)^{-\alpha+1} \quad (5)$$

with L_X in erg s⁻¹, R in cm, and E_s in keV. For convenience we have taken $\alpha = 1.6$ for each of the sources considered. This value represents a reasonable description of the observed spectra of Seyfert galaxies¹¹ and quasars¹². It is less precise for BL Lac objects, whose spectral variability makes the choice of an average α value difficult¹³. However, the function $f(\alpha = 1.6)$ only varies between 0.34 and 0.37 for the different energy bands over which the X-ray luminosities are calculated. Table 1 lists the values of L_X and $R = c \Delta t$ together with the optical depths τ for a range of γ -ray energies. Note that R represents an upper limit on the size of the X-ray region and consequently a lower limit on τ , because, in many cases, the Δt values are related to periods between observations rather than to measured changes in the X-ray emission.

Table 1 shows that Seyfert galaxies are transparent to their γ radiation up to energies of ~ 1 GeV. The size of the emitting region is probably not very different from one Seyfert to another so that the value of τ will ultimately depend on the X-ray luminosity of the object in question. Consequently, the less bright X-ray sources could be the best candidates for γ -ray emission. As Type II Seyferts have an average X-ray luminosity which is approximately two orders of magnitude lower than that of type I Seyferts^{14,15} we suggest that Type II objects should be primary targets for future γ -ray observations. In contrast it is clear that quasars and BL Lac objects must be effectively opaque to γ rays of a few MeV or more, if isotropic emission of both the X and γ rays is assumed. Many of the electron-positron pairs produced in this absorption process will eventually annihilate to produce 511 keV radiation, which is no longer absorbed by the X-ray photon field and can probably escape depending on the matter density present close to the emission region. If the positrons are slowed down such that the line is not smeared out by the relativistic kinematics, we may expect that these objects should be characterized by a weak 511 keV flux. Because in the case of the QSO 3C273 γ rays have been detected⁵ in the energy range 50–500 MeV, we must conclude that either the X- and γ -ray emission regions do not coincide or that anisotropic, highly collimated radiation is a characteristic of this active galactic nucleus. The degree of continuity in the spectral emission between the hard X-ray and the low energy γ -ray fluxes suggests that the possibility of beamed radiation is more likely. In the simplified model of jets consisting of streams of relativistic particles, equation (1) shows that as θ approaches zero, the threshold energy, E_s , tends towards infinity and the absorption process of γ rays becomes inhibited for higher and higher energies. For the case of a homogenous diverging beam with opening angle θ_0 , the limiting photon energy which is absorbed

Table 1 Photon-photon absorption for various extragalactic sources

Source	Type	$L_X(2-10 \text{ keV})$ (erg s^{-1})	Ref.	R (cm)	Ref.	Photon-photon optical depth as a function of γ -ray energy			
						1 MeV	10 MeV	10^2 MeV	10^3 MeV
NGC7469	S1	5.4×10^{43}	6	5.2×10^{15}	6	0.02	0.08	0.30	1.20
NGC4151	S1	6.2×10^{42}	22	$7.0 \times 10^{14*}$	22	0.02	0.06	0.26	1.03
NGC6814	S1	5.4×10^{42}	6	6.0×10^{14}	23	0.02	0.07	0.26	1.04
NGC3783	S1	2.3×10^{43}	6	5.2×10^{15}	11	0.01	0.03	0.13	0.51
MKN509	S1	2.5×10^{44}	24	3.1×10^{16}	24	0.01	0.06	0.23	0.94
MCG8-11-11	S1	8.3×10^{43}	25	7.8×10^{16}	25	0.002	0.008	0.03	0.12
NGC2992	S2	1.8×10^{43}	6	7.8×10^{15}	6	0.004	0.02	0.07	0.27
NGC5506	S2	1.3×10^{43}	26	7.8×10^{15}	26	0.003	0.01	0.05	0.19
NGC526A	S2	7.4×10^{43}	6	7.8×10^{15}	6	0.02	0.07	0.28	1.10
NGC7582	S2	5.3×10^{42}	6	5.7×10^{16}	27	1.7×10^{-4}	6.8×10^{-4}	0.027	0.01
MCG5-23-16	S2	2.2×10^{43}	28	1.0×10^{16}	28	0.004	0.02	0.06	0.26
3C273	QSO	1.1×10^{46}	6	$1.3 \times 10^{15+}$	6	15.6	61.9	246.5	981.2
2SO241+622	QSO	1.0×10^{45}	12	5.2×10^{15}	6	0.35	1.41	5.6	22.30
OX169	QSO	$1.1 \times 10^{44+}$	23	1.8×10^{14}	23	1.24	4.93	19.62	78.12
MKN421	BL Lac	2.2×10^{44}	6	2.6×10^{15}	6	0.16	0.62	2.46	9.81
PKS2155-304	BL Lac	$6.4 \times 10^{45}\S$	6	6.0×10^{14}	29	19.6	78.1	0.7	1,236.9
3G66A	BL Lac	$1.8 \times 10^{46}\ $	30	5.7×10^{15}	30	6.07	24.16	96.17	382.9

* X-ray variability also reported with $\Delta T \approx 700$ s (ref. 23) but of small statistical significance.

+ 10% X-ray variability also observed on $\Delta T = 610^3$ s (ref. 23).

‡ L_X in the range 0.5–4.5 keV.

§ Redshift measurement doubtful.

|| L_X in the range 0.2–3.5 keV.

by photon-photon pair production can be written as $E_\gamma \sim \sqrt{2} mc^2(1 - \cos \theta_0)^{-1/2}$. Thus to observe γ rays from 3C273 would require a beaming angle of few degrees or less. More realistic jet models^{16–20} involving relativistically moving plasmas, present a more complex picture. First the absorption calculation must be made in the rest frame of the flow where the emission is isotropic. Second, corrections must be applied for the combined relativistic effects of Doppler blueshifting of the γ -ray photons and of beaming which is responsible for increasing the source luminosity. The overall effect is to reduce the absorption of γ rays in emitters of beamed radiation.

It has been suggested^{19,20} that the sequence of extragalactic objects, radio-quiet quasars, radio-loud quasars and blazars (optically violent variable quasars and BL Lac objects) correspond to sources associated with relativistic jets, which are received at progressively smaller angles to their axes. In this picture, the nucleus may be considered to have an isotropic steady and unpolarized continuum which in the optical range is responsible for the emission lines and, in addition, a variable, strongly polarized source associated with a jet and beamed in a cone of small solid angle. The beamed component progressively dominates the isotropic emission as the angle between the observer and the jet axis decreases. Under this hypothesis, BL Lac objects probably represent the extreme case of alignment with the beam, so that the isotropic component and, consequently, the emission lines are completely swamped. Of the three quasars discussed here, two are classified as radio loud quasars²¹ and probably represent a less extreme case of alignment with the direction of the Earth. We would therefore expect both classes of objects to be γ -ray sources as along the beam direction little photon-photon absorption is expected. Conversely sources for which the beam is not directed towards the observer would not be detected as γ -ray emitters. If this is true, a future γ -ray survey of extragalactic objects should be carried out to test the possible correlation between the γ -ray spectral intensity and the angle of observation with respect to the jet axis. We conclude that γ -ray observations of extragalactic objects could not only test the beaming hypotheses but also support the argument that the classification of 'different' types of active galaxies is related to their orientation with respect to the Earth, rather than any intrinsic differences in their internal structure.

L.B. acknowledges the support of the British Council.

Received 8 June; accepted 8 October 1981.

1. Baity, W. A. *et al. Astrophys. J.* **244**, 429–435 (1981).
2. Perotti, F. *et al. Nature* **282**, 484–486 (1979).
3. Perotti, F. *et al. Astrophys. J.* **247**, L63–L66 (1981).

4. Perotti, F. *et al. Nature* **292**, 133–135 (1981).
5. Swannenburg, B. N. *et al. Nature* **275**, 298 (1978).
6. Marshall, N., Warwick, R. S. & Pounds, K. A. *Mon. Not. R. astr. Soc.* **194**, 987–1002 (1981).
7. Herterich, K. *Nature* **250**, 311–313 (1974).
8. Jelley, J. V. *Nature* **211**, 472–475 (1966).
9. Pollack, J. B., Guthrie, P. D. & Shen, B. S. P. *Astrophys. J. Lett.* **169**, L113–L116 (1971).
10. Jauch, J. M. & Rohrlich, F. *The Theory of Photons and Electrons* 299–301 (Springer, New York, 1976).
11. Mushotzky, R. F., Marshall, F. E., Boldt, E. A., Holt, S. S. & Serlemitsos, P. J. *Astrophys. J.* **235**, 377–385 (1980).
12. Worrall, D. M., Boldt, E. A., Holt, S. S. & Serlemitsos, P. J. *Astrophys. J.* **240**, 421–428 (1980).
13. Worrall, D. M., Boldt, E. A., Holt, S. S., Mushotzky, R. F. & Serlemitsos, P. J. *Astrophys. J.* **243**, 53–59 (1981).
14. Wilson, A. S. *Proc. R. Soc. A* **366**, 461–489 (1979).
15. Kriss, G. A., Canizares, C. R. & Ricker, G. R. *Astrophys. J.* **242**, 492–501 (1980).
16. Marscher, A. P. *Astrophys. J.* **235**, 386–391 (1980).
17. Konigl, A. *Astrophys. J.* **243**, 700–709 (1981).
18. Blandford, R. D. & Rees, M. J. *Phys. Scr.* **17**, 265–274 (1978).
19. Blandford, R. D. & Konigl, A. *Astrophys. J.* **232**, 34–48 (1979).
20. Scheuer, P. A. G. & Readhead, A. C. S. *Nature* **277**, 182–185 (1979).
21. Zamorani, G. *et al. Astrophys. J.* **245**, 357–374 (1981).
22. Lawrence, A. *Mon. Not. R. astr. Soc.* **192**, 83–94 (1980).
23. Tananbaum, H. X-ray Astronomy, Institute, Erice, Proc. NATO Advanced Study 291–310 (eds Giaconeri R. & Setti G.) (1979).
24. Dower, R. G., Griffiths, R. E., Bradt, H. V., Duxsey, R. E. & Johnston, M. D. *Astrophys. J.* **235**, 355–360 (1980).
25. Mushotzky, R. F. & Marshall, F. E. *Astrophys. J. Lett.* **239**, L5–L9 (1980).
26. Marshall, N. & Warwick, S. *Mon. Not. R. astr. Soc.* **189**, 37P–40P (1979).
27. Maccararo, T. & Perola, G. C. *Astrophys. J. Lett.* **246**, L11–L14 (1981).
28. Schnopper, H. W., Davis, M., Delvaile, J. P., Geller, M. J. & Huchra, J. P. *Nature* **275**, 719–721 (1978).
29. Snyder, W. A. *et al. Astrophys. J. Lett.* **237**, L11–L14 (1980).
30. Maccagni, D. & Tarengi, M. *Astrophys. J.* **243**, 42–46 (1981).

Are coronal loops stable?

I. J. D. Craig & A. N. McClymont

Department of Mathematics, University of Waikato, Private Bag, Hamilton, New Zealand

The solar corona is now recognized as a highly inhomogeneous plasma that comprises a complex network of individual arch and loop-like structures¹. This has led to the recent idea that the isolated 'coronal loop' may form a basic building block of quiescent solar and stellar atmospheres^{2,3}. Theoretical studies^{4,5}, however, suggest that isolated loops are violently unstable and hence should collapse on the radiative time scale of the plasma. The simple fact that loops exist stably over long periods demonstrates the inadequacy of the current theory. We point out here that a crucial ingredient is missing from the theory and hence reconcile the existence of coronal loops with the disruptive effect of the radiative instability.

The central assumption in theoretical loop modelling is that dynamic effects can be neglected in the local energy balance⁶.

Hence it is postulated that an unknown heating mechanism supports the loop against energy losses by radiation and energy redistribution by heat conduction along the confining magnetic field lines. This leads quite naturally to a scaling law which, for the loop of length l , relates the plasma pressure P to the peak coronal temperature in the loop:

$$P \approx 2.1 \times 10^{-11} \beta^{1/2} T_0^{13/4} / l \quad \text{dyn cm}^{-2} \quad (1)$$

where β is a number of the order of unity accounting for the shape of the heating function⁷. An important prediction of the quasi-static model therefore is that small, hot loops should be several orders of magnitude denser than larger, cooler structures. That such behaviour is confirmed observationally is very strong support for the quasi-static analysis.

Yet an astrophysical plasma hotter than 10^5 K may be unstable because the radiative loss rate decreases with increasing temperature. Field's stability criterion⁷ for a uniform conducting plasma is

$$P < 6.1 \times 10^{-11} T^{13/4} / l \quad \text{dyn cm}^{-2} \quad (2)$$

Although equation (2) is not strictly applicable to a non-uniform coronal loop, comparing equations (1) and (2) we anticipate that conduction may not be able to control the instability. Independent normal mode stability analyses in fact claim that the 'thermally isolated' loop (that is, a loop whose temperature gradient vanishes at the lower boundary) is radiatively unstable^{4,5}. This result implies that if the loop is to exist at all it cannot be described by a quasi-static energy balance. Thus it has been speculated that a dynamic solution may exist in which the radiative instability drives periodic dynamic oscillations on the radiative time scale of the plasma^{4,6}.

This conclusion is disturbing for two main reasons: first, quiescent loops show little evidence of significant dynamical behaviour, nor are they observed to collapse on the radiative time scale of the plasma¹⁻³. Second, detailed hydrodynamic models based on the same physics as the quasi-static analysis do not reveal oscillatory modes driven by the radiative instability⁸⁻¹⁰.

In attempting to resolve this problem we have applied a standard normal mode analysis⁴ to a 'thermally isolated' loop (see also refs 5, 11). Two classes of lower boundary conditions are considered: in one case the base is held at a fixed chromospheric temperature ($\delta T_b = 0$); in the other the heat flux is assumed to vanish at the base ($\delta F_b = 0$). As the former condition implies that the chromosphere can act as an infinite source or sink of heat while the latter implies that no heat is exchanged across the lower boundary⁹, we expect that a real loop must lie between these extremes. In either case, because the base of the loop cannot move freely¹² ($\delta v_b = 0$, where v is the velocity), the most unstable eigenmode is the first antisymmetric mode⁴—symmetric modes tend to be stabilized by the rise in pressure which accompanies a temperature increase.

It can be proved that the loop is marginally stable in the case where the base temperature is held fixed. To see this, we note that the first antisymmetric eigenfunction given by

$$\delta T \propto T', \quad \delta v \propto T - T_b \quad (3)$$

where T' denotes the unperturbed temperature gradient, represents an analytic solution to the problem⁴ corresponding to the case of zero growth. This result also establishes that the loop is unstable when $\delta F_b = 0$, as, by Sturm's oscillation theorem⁵, the eigenvalue λ (the negative of a normalized growth rate⁴) must decrease to satisfy the required boundary condition. However, the instability is physically insignificant if the amount of radiatively stable chromospheric material ($T < 10^5$ K) at the foot of the transition region is sufficiently large. The growth time is then greater than any relevant time scale associated with the loop.

Figure 1 shows the first antisymmetric ($\delta F_b = 0$) eigenvalue λ plotted against the ratio of chromospheric ($T < 10^5$ K) mass to coronal ($T > 10^5$ K) mass in the loop, N_{CH}/N_{CO} , for various base temperatures T_b . A slight increase in chromospheric mass clearly produces a dramatic decrease in $|\lambda|$; likewise the growth rate of the instability decreases. From the definition⁴ of λ and

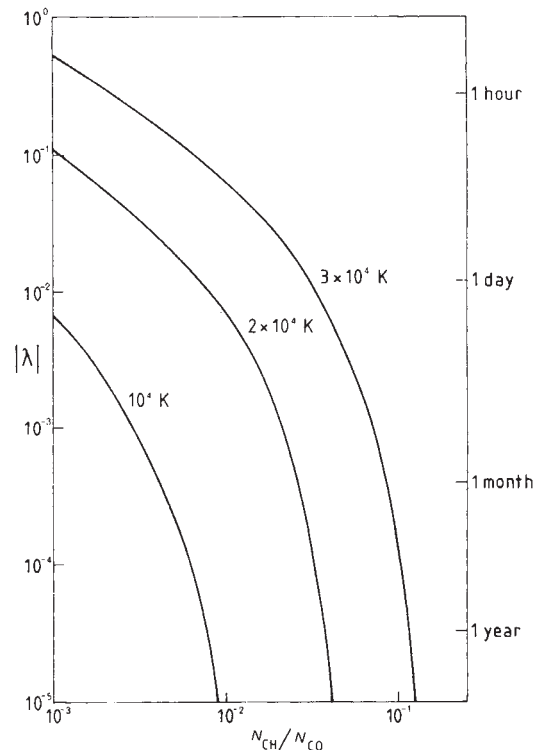


Fig. 1 The first antisymmetric eigenvalue λ (normalized growth rate) plotted against chromospheric mass ($T < 10^5$ K) for a range of base temperatures T_b . The amount of chromospheric material N_{CH} is expressed in units of the coronal mass N_{CO} . On the right-hand scale actual growth times for the typical values $T_0 = 2.5 \times 10^6$ K and $l = 10^{9.5}$ cm are shown.

the scaling law (1) we find (setting $\beta = 1$) the actual growth time

$$\tau = 2.7 \times 10^{-5} \frac{l}{|\lambda| T_0^{1/4}} \quad \text{s} \quad (4)$$

which for the typical values $T_0 = 2.5 \times 10^6$ K and $l = 10^{9.5}$ cm yields the growth times shown on Fig. 1. Values of τ greater than a few weeks have no physical significance as the growth time exceeds the lifetime of a typical loop.

We conclude that the radiative instability is controlled essentially by the chromosphere: as more and more of the radiatively stable chromospheric plasma is included in the analysis the ($\delta F_b = 0$) mode of instability tends towards the marginally stable case ($\delta T_b = 0$). Very little chromospheric material is, in fact, required to counter the instability: between 1 and 10% of the coronal mass suffices. It follows that the rapid instabilities found previously^{4,5,11} are simply a consequence of the neglect of the chromospheric influence.

Finally, we note that a complete solution to the problem requires the treatment of energy transport by radiative transfer in the chromosphere. Because only a minute fraction of the radiatively stable material available is required to quash the instability in the optically-thin approximation used above, we believe our conclusions will remain essentially unaltered by the inclusion of radiative transfer.

Received 19 August; accepted 8 October 1981.

- Orral, F. Q. (ed.) *The Skylab Workshop on Active Regions* (Colorado Associated University Press, 1981).
- Craig, I. J. D., McClymont, A. N. & Underwood, J. H. *Astr. Astrophys.* **70**, 1-11 (1978).
- Rosner, R., Tucker, W. H. & Vaiana, G. S. *Astrophys. J.* **220**, 643-655 (1978).
- Antiochos, S. K. *Astrophys. J.* **232**, L125-L130 (1979).
- Hood, A. W. & Priest, E. R. *Astr. Astrophys.* **87**, 126-131 (1979).
- Craig, I. J. D. *Solar Flare MHD* (ed. Priest, E. R.) Ch. 5 (Gordon & Breach, London, 1981).
- Field, G. B. *Astrophys. J.* **142**, 531-555 (1965).
- Craig, I. J. D. & McClymont, A. N. *Solar Phys.* **70**, 97-113 (1981).
- Craig, I. J. D., Robb, T. D. & Rollo, M. D. *Solar Phys.* (in the press).
- McClymont, A. N. & Canfield, R. C. *IAU Symp. No. 41*, 313-316 (1980).
- Priest, E. R. *The Skylab Workshop on Active Regions* Ch. 9 (ed. Orall, F. Q.) (Colorado Associated University Press, 1981).
- McClymont, A. N. & Craig, I. J. D. *University of Waikato Mathematics Res. Rep. No. 96* (1981).

Dayside electron cyclotron harmonic emissions

Kjell Rönmark

Kiruna Geophysical Institute, University of Umeå,
S-901 87 Umeå, Sweden

Peter J. Christiansen

Space and Plasma Physics Group, University of Sussex,
Brighton BN1 9QH, UK

Observations of waves exhibiting an often complex frequency banding related to the electron gyrofrequency were first reported in 1970¹. Later observations²⁻⁵ showed that the spectral content of these electrostatic electron cyclotron harmonic waves is highly variable^{6,7}, including single frequency emissions between the electron gyrofrequency Ω_e and its first harmonic, multiple bands between consecutive harmonics of Ω_e , and emissions in a single band far above the electron gyrofrequency⁸⁻¹⁰. These observations have led theorists to investigate both the linear instability which generates the waves¹¹⁻¹⁵ and various wave-particle^{14,16} and wave-wave¹⁷⁻¹⁸ interactions in which they may be involved. The consensus seems to be that at least two plasma components with different temperatures are required to destabilize electron cyclotron harmonic waves. We adopt this view here and test our present knowledge of electron cyclotron harmonic instabilities against the observed reality. To do this we use simultaneous observations of wave spectra and electron distribution functions, available in the GEOS 1 data from 25 August 1977 (see ref. 15). From these we present the wave event and then use the measured particle fluxes to derive a model distribution function, providing the input to a computer program which solves the plasma dispersion relation. From the computed temporal growth rates and group velocities, the total amplification of waves which are unstable within a limited volume of space is estimated.

The emissions observed by the S300 wave experiment onboard GEOS 1 between 09.20 and 10.20 UT on 25 August 1977 are shown in Fig. 1. The spectrogram is derived from the large electric antenna¹⁹ and consists of sets of 0–77 kHz sweeps of the stepped frequency analyser. The grey scaling is adjusted to the strongest signal in each sweep. The horizontal lines at 47.6 and 63.5 kHz are caused by the satellite clock pulses and are thus purely instrumental. The line f_{ce} indicates the electron gyrofrequency, which during the event was ~ 4.4 kHz. The variable high frequency emissions seen before 09.40 UT are known to be in the vicinity of the upper hybrid (or plasma) frequency. When the plasma frequency drops to around 30 kHz after 09.40 UT, signals appear between all gyroharmonics below the intensified plasma line. The signal at the second harmonic of the plasma frequency is believed to be (at least partly) instrumental, but it indicates a very strong signal at the fundamental. The emissions observed after 09.40 UT would belong to class 3 in the classification schemes of Gough *et al.*⁷ and of Hubbard and Birmingham⁶. The higher harmonics disappear at around 09.53 UT, while the $3/2 \Omega_e$ emission persists until 10.00 UT. After this time, only the upper hybrid (or plasma) line remains. At 09.50 UT, the satellite was on the dayside (11.20 MLT) at $L \sim 6.6$ and $\sim 10^\circ$ magnetic latitude.

The importance of the cool plasma parameters for electron cyclotron harmonic instabilities is discussed in the accompanying paper by Horne *et al.*²⁰. Low energy (0–500 eV) electron fluxes are measured by the S302 experiment onboard GEOS 1 (ref. 21), and information about the low energy electrons at 09.47 UT has been provided by J. Johnson, J. Sojka and G. Wrenn (personal communication). The main components have temperatures $T_1 \approx 1$ eV and $T_2 \approx 5$ eV, and the densities are

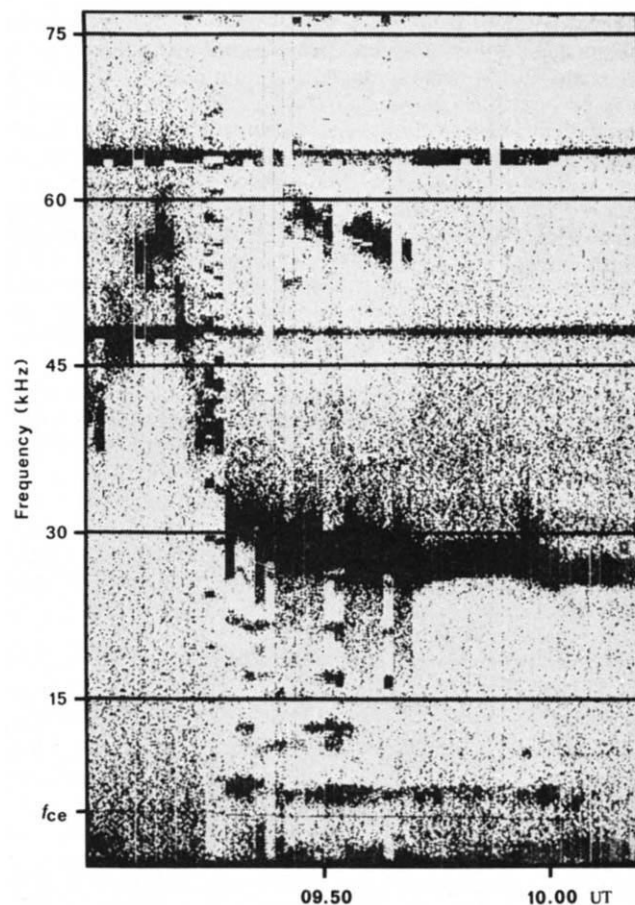


Fig. 1 Spectrogram emphasizing the wave event under study. The display begins at 09.20 UT the time scale is not uniform.

$n_1 \approx 5.5 \times 10^6$ electron m^{-3} and $n_2 \approx 3.0 \times 10^6$ electron m^{-3} . In addition, a component with temperature $T_3 \approx 200$ eV and density 5×10^4 electron m^{-3} is observed. The flux of this component at a pitch angle of 80° is double the flux at 18° . This anisotropy and the slope of the energy spectrum match the low energy end of the distribution measured by the S310 experiment, although the flux values given by S310 are higher.

The particle experiment S310 provides pitch angle distributions and energy spectra in the range 0.35–22 keV (ref. 22). When the angle between the magnetic field and the spin axis is 15° – 26° , at least one of the four detectors with a narrow field of view ($\pm 2^\circ$) passes through the loss-cone. In such cases even the very narrow loss-cone expected at geostationary altitudes can be resolved. At larger pitch angles, measurements are made by

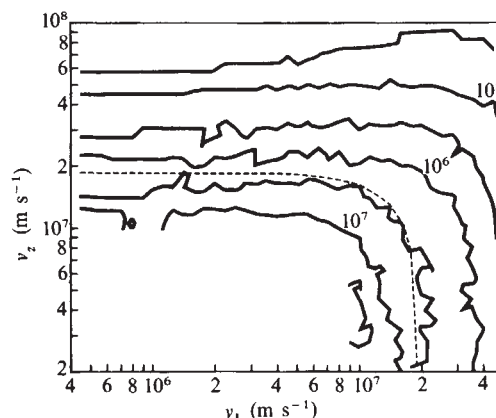


Fig. 2 The electron distribution function observed by S-310 at 09.45–09.48 UT. Phase space densities are given in units of electron $cm^{-2} s^{-1} sr^{-1} keV^{-1}$, and the dashed curve indicates 1 keV.

three other detectors. All seven detectors step through 32 energy steps, and a complete distribution function such as the one shown in Fig. 2 is obtained in ~ 3 min. As each detector is sampled 11.6 times per second, the distribution shown in Fig. 2 would ideally be based on $\sim 15,000$ samples, but as some data contaminated by sunlight and other effects must be rejected, the real number may be a few thousands lower. The pitch angle intervals scanned by different detectors normally overlap, and this overlap is used by an intercalibration routine to compensate for differences in the efficiency of the electron multipliers.

The contours of constant level plotted in Fig. 2 are derived by dividing the measured fluxes by the particle energy and transforming the energy and pitch angle of each sample to $(v_\perp, v_z)$ -coordinates. The displayed part of velocity space is then covered by a network, consisting of 43×35 meshes. Between each mesh point, the logarithm of a velocity component changes by 0.05. The average value of the distribution function is calculated within each mesh, and linear interpolation between nearest neighbours is used to define the function between the mesh points. As the pitch angle resolution is limited to 2° and the logarithmic scales will expand the small pitch angle region strongly, the average flux in the intervals 0° – 2° and 2° – 4° pitch angle has been used. Insignificant ripple of the contours at small pitch angles can thus be avoided.

The dashed curve at 1 keV shows the shape of an isotropic distribution. The value of the distribution function is lower by a factor of ~ 3 at small pitch angles than at $\alpha = 90$, but the positive slope in the distribution is apparently very weak.

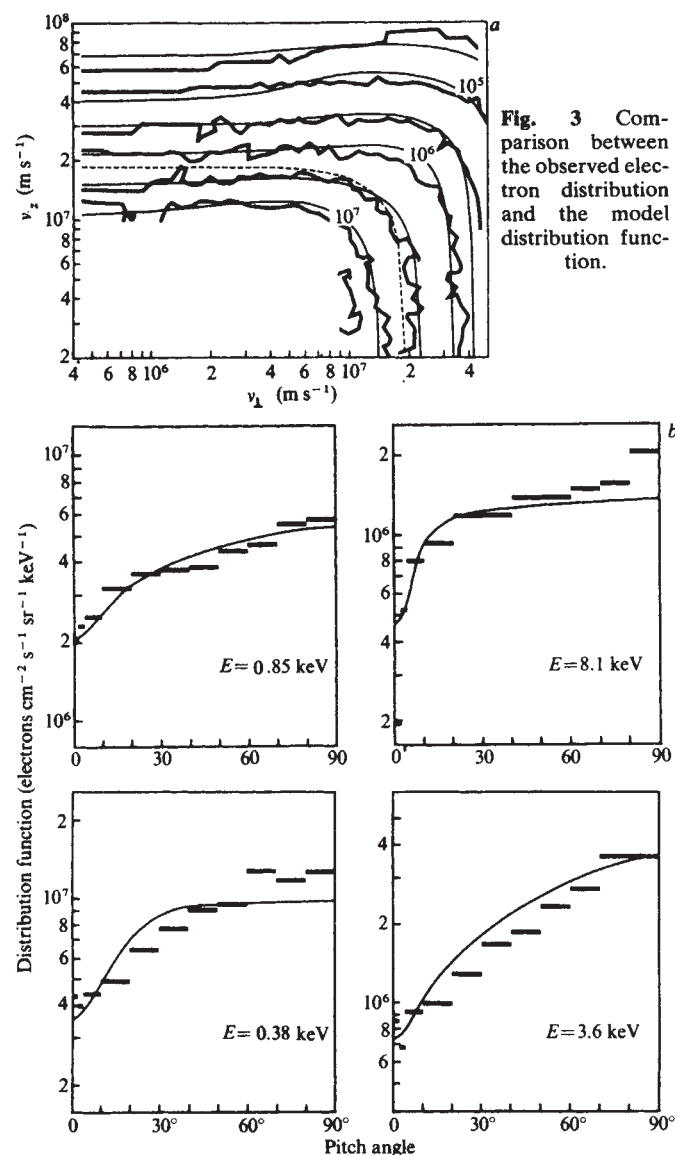


Fig. 3 Comparison between the observed electron distribution and the model distribution function.

Table 1 Parameters of the complete model distribution

Component no.	Density (electron m ⁻³)	Temperature (eV)	Δ	β
1	5.5×10^6	1	1	—
2	3.0×10^6	5	1	—
3	6.6×10^4	200	0.35	0.15
4	6.4×10^4	1,000	0.35	0.99
5	5.1×10^4	8,000	0.40	0.02

A computer code can be used to estimate the growth rates of possible instabilities, if the measured distribution is modelled by a sum of components of the form¹²

$$f_i = \frac{\exp\left(-\frac{v_z^2}{V_i^2}\right)}{(\pi V_i^2)^{3/2}} \left\{ \Delta_i \exp\left(-\frac{v_\perp^2}{V_i^2}\right) + \frac{1-\Delta_i}{1-\beta_i} \times \left[\exp\left(-\frac{v_\perp^2}{V_i^2}\right) - \exp\left(-\frac{v_\perp^2}{\beta_i V_i^2}\right) \right] \right\}$$

Here, the parameters Δ_i and β_i measure the depth and width of the loss-cone, and $V_i = (2T/m)^{1/2}$ is the thermal velocity. A model of this kind fitted to the observations is shown in Fig. 3. Figure 3a shows the contours of the model superimposed on the measured distribution. The differential fluxes as a function of α are compared in Fig. 3b. It is not easy to assess the degree of agreement between model and measurement quantitatively, as the fit is done by trial and error. The model shown in Fig. 3 is close to the best obtainable using three components of the form above.

In addition to the three hot components, two maxwellian components are used to model the low-energy electrons. The parameters of the complete model distribution are given in Table 1.

The wave spectra have been calculated numerically by means of a computer code which solves the dispersion relation of waves in a magnetized plasma consisting of maxwellian components. The dielectric tensor of a maxwellian plasma can be expressed in terms of modified Bessel functions I_n , and the plasma dispersion function Z (ref. 23). In the program, the dielectric tensor and its first derivatives with respect to ω , k_z and $k_\perp$ are written as combinations of I_n and $Z(\omega - n\Omega_e)/k_z$. The full electromagnetic dispersion relation is then solved by iteration, using Newton-Raphson's method. Any two of the three variables ω , k_z and $k_\perp$ can be specified and solved for the third, but because $\mathbf{k}$ is taken to be real while ω is complex, convergence can be ensured only when solving for ω , and only this case will be considered here. The iteration stops when the relative error in the last estimate is $< 10^{-6}$. We have then obtained a real frequency ω_k and a temporal growth rate γ_k . However, to estimate the wave spectrum we need also the group velocity v_g . The components of the group velocity are easily calculated from the derivatives of the dispersion relation, and the intensity amplification of the wave can be estimated as γ_k/v_g times the propagated distance.

When the wave propagates a distance l , the gyrofrequency will change by $\Delta\Omega_e \approx l \cdot \nabla\Omega_e$. As the unstable bandwidth is $\sim 0.2 \Omega_e$ (in all harmonic bands), the amplification will stop when

$$\left| \frac{\omega}{\Omega_e} - \frac{\omega}{\Omega_e + \Delta\Omega_e} \right| \approx 0.2$$

In a dipole field we have $\Delta\Omega_e/\Omega_e \approx 3l/r_0$, where l is the distance propagated in the radial direction, and r_0 is the distance from the Earth. This leads to

$$L_r \approx \frac{0.2\Omega_e}{3\omega} r_0$$

as the maximum radial distance over which the wave can be amplified. At $r_0 \approx 6 R_E$ and $\omega \approx 3/2 \Omega_e$, this corresponds to $\sim R_E/4$. However, the radial amplification length may not be very interesting, as the scale lengths parallel to the magnetic field and in azimuth are larger.

Experimental studies of the occurrence of these types of emission indicate that they are confined to within a few degrees of the geomagnetic equator^{7,10}. Although the reasons for this close confinement to the equatorial plane are poorly understood, a maximum amplification length $L_z = 2 L_r$ has been used in the calculations. This corresponds to a confinement within $\sim \pm 2.5^\circ$ of the geomagnetic equator. There is no way of determining the maximum amplification length $L_\perp$ in azimuth from observations. As a reasonable estimate we chose $L_\perp = 4 L_r$. These estimates are very rough, but an accurate determination of the amplification lengths would require ray tracing in a magnetosphere where the variation of the plasma parameters in a large volume of space were known. As there is no experimental technique which measures the relevant parameters throughout a large volume, we must rely on such estimates, but note that the value of $L_\perp$ chosen is compatible with the result of a simplified ray tracing in an idealized magnetospheric model¹².

Assuming that the wave growth is limited by propagation effects rather than some nonlinear saturation mechanism, we can calculate the shape of the frequency spectrum. At each frequency, we choose the k value which gives the strongest amplification within a region of space with dimensions $L_\perp$ and L_z . The amplification (in dB) is calculated as $8.7 \gamma_k \cdot \min(L_\perp/v_{g\perp}, L_z/v_{gz})$ and this amplification rate as function of ω_k will give the approximate shape of the frequency spectrum. Absolute values of the spectral density cannot be calculated without specifying the initial level of fluctuations, but the relative strength of emissions at different frequencies can be compared with observed spectra.

The amplification predicted by the theory is compared in Fig. 4 with a measured frequency spectrum. The spectral density

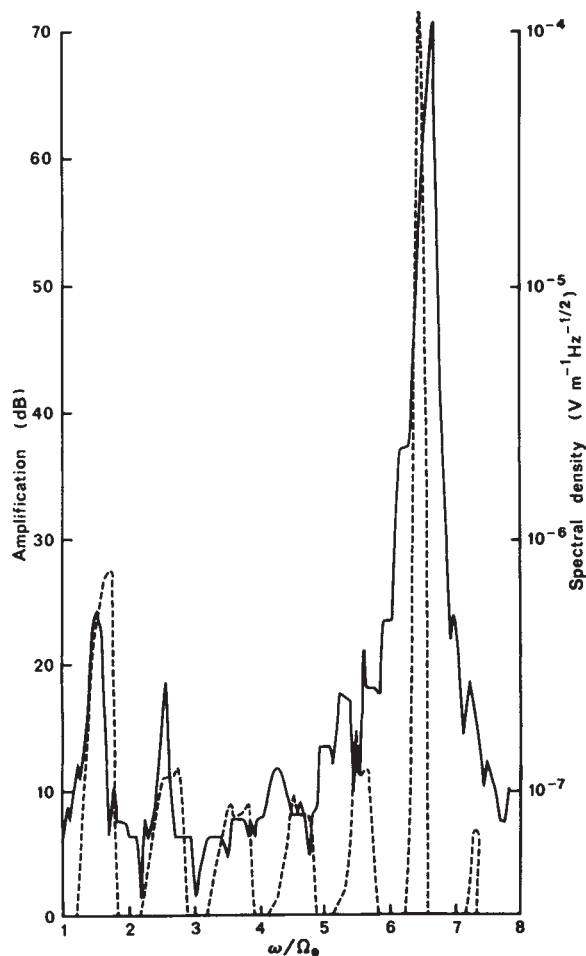


Fig. 4 Comparison between the spectral density observed by S-300 at 09.48 UT and the calculated amplification (dashed).

(solid line) was measured by the S300 experiment at around 09.48 UT on 25 August 1977. The density amplification (a dashed line) has been calculated using the model distribution function defined in Table 1. Although the exact positions of the peaks within the harmonic bands coincide poorly, the general shape of the measured spectrum is very well described by the calculated amplification. The theory correctly predicts that the strongest emission is near the upper hybrid frequency, with fairly high amplitudes around $3/2 \Omega_e$ and weaker signals in the intermediate harmonic bands. The upper hybrid amplitude of the very quietest emissions observed by GEOS 1 (ref. 9) is $\sim 5 \times 10^{-8} \text{ V m}^{-1} \text{ Hz}^{-1/2}$. Taking this value as a reasonable estimate of the background fluctuation level, we find that it is consistent with the 0 dB level chosen for the comparison in Fig. 4.

Two main types of error will affect the reliability of this comparison of theory and experiment, the most important source probably being the limited time resolution of the experiments. A measurement of the complete hot particle distribution cannot be made in < 3 min, in which time all relevant parameters may vary significantly. As the low energy (0.3–1.5 keV) particles mainly involved in the instability were measured at the end of the period 09.45–09.48 UT, the spectrum chosen for the comparison has been taken near 09.48 UT to minimize the effect of time variations. However, a full frequency sweep is obtained in 22 s, and significant changes in the spectrum are often seen from one sweep to the next. This indicates that rapid variations are common, and the accuracy of the prediction is probably limited by these time variations. The second source of error is the rough estimate of the size of the amplification region. If the magnitudes of $L_\perp$ and L_z are varied with the ratio $L_\perp/L_z$ fixed, the maximum amplification at all frequencies will vary at the same rate. The amplification curve will be compressed or expanded but retain its general shape. Reasonably small changes in the ratio $L_\perp/L_z$ should mainly influence the frequency of maximum amplification within each harmonic band, but small changes in the magnitude of the peaks might also occur. Compared with these two main difficulties, the errors introduced when modelling the distribution function and solving the dispersion relation are probably small.

The values of most parameters needed for this study have been determined directly from measurements, and the remaining values are at least consistent with observations. Given a set of input parameter values, it seems unlikely that the calculations would reproduce the observed wave spectrum so well if the essential physics were not included in the mathematical model. In particular, it seems possible to estimate the shape of the wave spectrum using linear theory if the particle distribution function is known. Quasilinear or nonlinear effects are apparently not required to limit the wave growth in this case, although the weakness of the gradients in the hot electron distribution may well be the result of quasilinear diffusion.

Considering the present results and those of Horne *et al.*²⁰, our understanding of the linear theory of electrostatic electron cyclotron harmonic waves is probably approaching a point where quantitative predictions about wave spectra may be based solely on particle measurements. The experience of applying this methodology to a well chosen set of wave events should also enhance the possibilities of using wave measurements for magnetospheric plasma diagnostics.

This study has been performed within a large GEOS 1 collaboration and our colleagues have contributed to it in many ways. In particular we thank H. Borg, P. Gough and R. Horne for assistance in the data handling and valuable comments. This research was supported by the Swedish Board for Space Activities, the SRC and the ESA.

Received 14 November 1980; accepted 26 August 1981.

1. Kennel, C. F., Scarf, F. L., Fredricks, R. W., McGehee, J. G. & Coroniti, F. V. *J. geophys. Res.* **75**, 6136 (1970).
2. Fredricks, R. W. & Scarf, F. L. *J. geophys. Res.* **78**, 310 (1973).
3. Scarf, F. L., Fredricks, R. W., Kennel, C. F. & Coroniti, F. V. *J. geophys. Res.* **78**, 3119 (1973).

4. Shaw, R. R. & Gurnett, D. A. *J. geophys. Res.* **80**, 4259 (1975).
5. Anderson, R. R. & Maeda, K. *J. geophys. Res.* **82**, 135 (1977).
6. Hubbard, R. F., Birmingham, T. H. & Hones, E. W. *J. geophys. Res.* **84**, 5828 (1979).
7. Gough, M. P., Christiansen, P. J., Martelli, G. & Gershuny, E. J. *Nature* **279**, 515 (1979).
8. Kurth, W. S., Craven, J. D., Frank, L. A. & Gurnett, D. A. *J. geophys. Res.* **84**, 4145 (1979).
9. Christiansen, P. *et al. Nature* **272**, 682 (1978).
10. Christiansen, P. *et al. Space Sci. Rev.* **22**, 383 (1978).
11. Young, T. S. T., Callen, J. D. & McCune, J. E. *J. geophys. Res.* **78**, 1082 (1973).
12. Ashour-Abdalla, M. & Kennel, C. F. in *Physics of the Hot Plasma in the Magnetosphere* (eds Hultqvist, B. & Stenflo, L.) 201 (Plenum, New York, 1975).
13. Karpman, V. I., Alekhin, Ju. K., Borisov, N. D. & Rajabova, N. A. *Plasma Phys.* **17**, 937 (1975).
14. Ashour-Abdalla, M. & Kennel, C. F. *J. geophys. Res.* **83**, 1531 (1978).
15. Rönnmark, K., Borg, H., Christiansen, P. J., Gough, M. P. & Jones, D. *Space Sci. Rev.* **22**, 401 (1978).
16. Lyons, R. L. *J. geophys. Res.* **79**, 575 (1974).
17. Rönnmark, K. *Planet Space Sci.* **25**, 149 (1977).
18. Melrose, D. B. *Theoretical Physics Preprint* (University of Sydney, 1980).
19. Jones, D. *Space Sci. Rev.* **22**, 327 (1978).
20. Horne, R. B. *et al. Nature* **294**, 338–340 (1981).
21. Johnson, J. F. E., Sojka, J. J. & Wrenn, G. L. *Space Sci. Rev.* **22**, 567 (1978).
22. Borg, H. *et al. Space Sci. Rev.* **22**, 511 (1978).
23. Fried, B. D. & Conte, S. D. *The Plasma Dispersion Function* (Academic, New York, 1961).

Amplitude variations of electron cyclotron harmonic waves

R. B. Horne*, P. J. Christiansen*, M. P. Gough*, K. G. Rönnmark†, J. F. E. Johnson‡, J. Sojka‡ & G. L. Wrenn‡

* Space and Plasma Physics Group, University of Sussex, Brighton BN1 9QH, UK

† Kiruna Geophysical Institute, University of Umeå, S-901 87 Umeå, Sweden

‡ Mullard Space Science Laboratory, University College London, Gower Street, London WC1E 6BT, UK

Electron cyclotron harmonic (ECH) instabilities just outside the plasmopause and at frequencies near the cold upper hybrid frequency are a common feature of the Earth's magnetosphere. These waves which have virtually no magnetic component, are believed to have an important role in the generation of weak diffuse aurora^{1,2}. They are able to interact strongly with electrons in the hundred eV to several keV energy range which can result in pitch angle scattering and precipitation on magnetic field lines which map down into the auroral zone. On the dayside magnetosphere these waves can exhibit large amplitude variations of 30–40 dB and can also exist at steady amplitudes on time scales of the order of tens of seconds. Here we seek an explanation for the sporadic nature of such instabilities by performing linear stability calculations and extending the technique used in the accompanying paper³.

An example of the emissions recorded by GEOS 1 on 25 August 1977 is shown in the frequency time spectrogram of ref. 3. Figure 1 of ref. 3 shows that after 09.45 UT in that event, the signal in the band near the upper hybrid frequency f_{uh} is accompanied by lower-frequency emissions at $\sim 3/2$ and $\sim 5/2$ times the electron gyrofrequency f_{ce} typical of spectra of class 3 in the scheme of Gough *et al.*⁴.

In the periods 09.48–09.50, 09.51–09.52 and 09.53–09.54 UT the second harmonic of the strongest emission appears together with signals in the extra low frequency (ELF) range. The second harmonic signal is an instrumental saturation effect indicative of large wave amplitudes $> 3 \text{ mV m}^{-1}$. The ELF signal measured in the frequency range 15–450 Hz and detected by the same preamplifier also saturates and thus preamplifier saturation effects cannot be ruled out. However, studies of similar events on GEOS 1⁵, have shown that ELF signals detected in unsaturated conditions are electrostatic with a spectral peak close to the lower hybrid frequency. Nonlinear wave-wave couplings due to the strong f_{uh} pump wave are thus implied.

Time variation of f_{uh} and ELF amplitudes are shown in Fig. 1, with changes in the energy density of electrons below 500 eV.

Electrons in the energy range 0.5–500 eV are measured by the S302 particle experiment⁶ using two detectors, one looking along the satellite spin axis and the other at 80° to it. As the angle between the spin axis and the magnetic field was 18° during this event, pitch angles between 62° and 98° were covered by the second detector.

Analysis of the data from this instrument shows, after subtraction of photoelectron contamination, that the low energy electrons can be represented by four maxwellian components at temperatures of 1, 5, 60 and 200 eV. Only the 5-eV component shows sporadic variations which are seen preferentially in the nearly field-aligned detector. Summarizing the observations in terms of an energy density, this quantity is strongly anticorrelated with the strength of waves near f_{uh} and ELF emissions as shown in Fig. 1. This suggests that the wave amplitude variations might be connected with changes in the distribution of low-energy electrons.

Higher-energy electron fluxes between 350 and 22 keV, recorded by the S310 experiment⁷, show remarkably little variation by comparison. Data taken at intervals of 6 min, and covering times of highly variable wave activity, reveal no significant changes in the hot electron distribution. The distribution shown in Fig. 2 of ref. 3 can thus be regarded as typical for the time period of interest here.

Using the techniques described in ref. 3, two electron distributions have been constructed, one (model A) appropriate to the period of strong wave activity 09.48–09.50 UT and the other

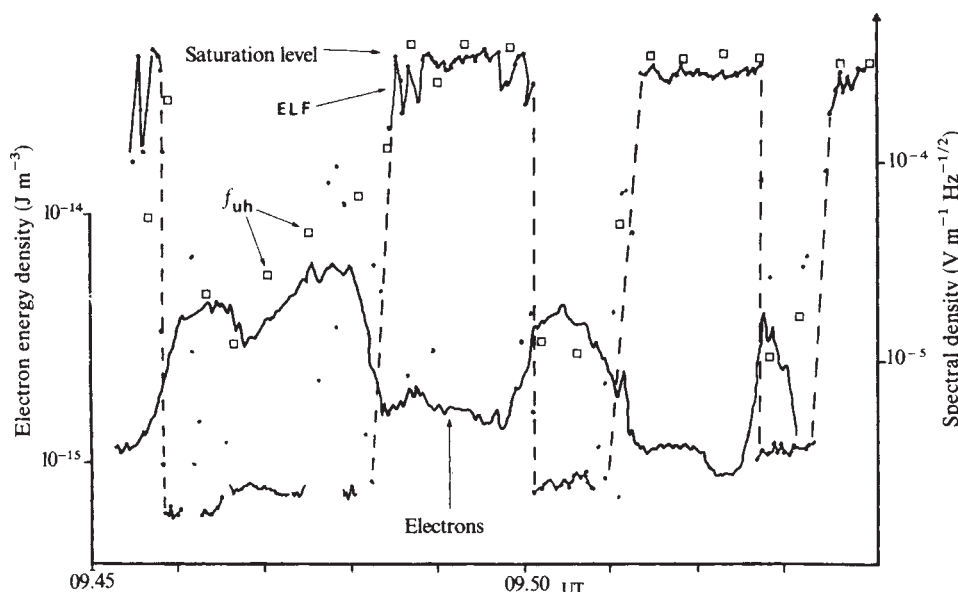


Fig. 1 Anticorrelation between the energy density of electrons below 500 eV and the amplitude of waves both in the ELF region and near the upper hybrid frequency f_{uh} . A full pitch angle energy distribution is measured by the S310 experiment between 09.45 and 09.48 UT. At 09.50 UT the satellite coordinates are LT = 11.20, $L = 6.6$, magnetic latitude = 10° and $k_p = 5$.

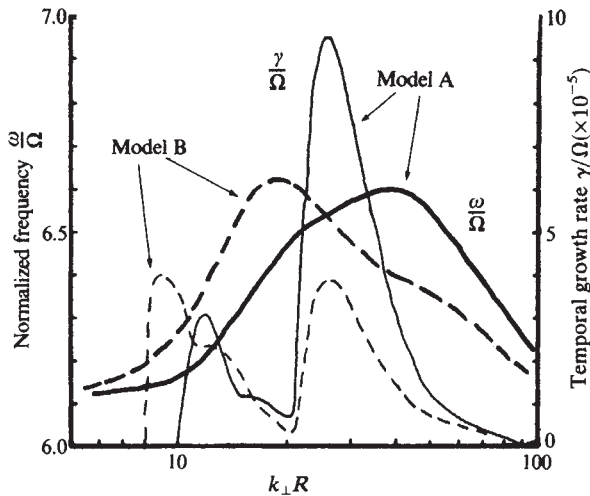


Fig. 2 Solution of the dispersion relation between the electron gyroharmonics 6Ω and 7Ω . The electron gyrofrequency is $f_{ce} = 4.4$ kHz and the cold upper hybrid frequency is $f_{uh} = 6.05f_{ce}$. The thick solid and dashed curves are the wave frequency for each model while the thin sharply peaked curves are the temporal growth rates. Calculated parameters include $\alpha = 52$, $\Delta\omega = 0.1$ for model A and $\alpha = -22$, $\Delta\omega = 0.2$ for model B.

(model B) corresponding to weaker wave activity 09.45–09.48 UT. Both models use the same parameters derived in ref. 3 for the three highest energy components, while the low energy components have been remodelled from a more refined analysis of the low-energy particle data: model A consists of two isotropic maxwellians with temperatures and densities, $T_1 = 1$ eV, $n_1 = 7.5 \times 10^6 \text{ m}^{-3}$, $T_2 = 5$ eV, $n_2 = 1 \times 10^6 \text{ m}^{-3}$. For the three components of model B the following values are used: $T_1 = 1$ eV, $n_1 = 4.5 \times 10^6 \text{ m}^{-3}$, $T_2 = 5$ eV, $n_2 = 3.5 \times 10^6 \text{ m}^{-3}$; $T_{3z} = 10$ eV, $T_{3\perp} = 5$ eV, $n_3 = 5 \times 10^5 \text{ m}^{-3}$. This last anisotropic component, plus the increased density of n_2 , reflect reasonably well the increased fluxes at small pitch angles of suprathermal electrons which are presumably of ionospheric origin. The cold upper hybrid frequency is set at 26.6 kHz, a little lower than in previous preliminary reports^{8,9}.

The linear electrostatic dispersion relation for ECH wave propagation in a multi-component plasma has been solved numerically for complex frequency as a function of real wave vector components, $k_\perp$, k_z are respectively perpendicular and parallel to the ambient magnetic field $\mathbf{B}_0$. The dependence of the normalized real frequency ω/Ω (where $\Omega = 2\pi f_{ce}$) is shown in Fig. 2 as a function of $k_\perp R$ (R being the 200-eV electron Larmor radius of ~ 300 m). Also plotted are the largest normalized temporal growth rates γ/Ω taken for the unstable k_z range.

A comparison of the two models shows that the enhanced 5-eV component in model B has moved the frequency maximum to smaller perpendicular wavenumbers and increased the perpendicular group velocity near $k_\perp R \sim 30$, where γ is large. Furthermore, increased cyclotron damping, caused by the additional 5-eV electrons, has reduced the temporal growth rate by a factor of 2 with respect to that of model A. However, the unstable waves predicted here have non-zero group velocities. They are convectively unstable and therefore the spatial growth rates must be considered to determine the final wave amplification levels¹⁰.

Spatial growth rates k_i can be calculated from the temporal growth rate and the group velocity ($k_i = -\gamma/V_g$, $k_i \ll k_\perp$). The final amplification levels are estimated when the distance over which the amplifying waves propagate is known.

It is assumed that the plasma density variation is small over the relevant time scales and distances for convective growth.

Furthermore, changes in the dimensionless frequency $\bar{\omega} = \frac{\omega}{\Omega}$ are assumed only to occur as the waves propagate through regions of

varying magnetic field (assumed to be dipolar). When the change in dimensionless frequency, $\Delta\bar{\omega}$, is such that the waves are effectively detuned from resonance, convective growth ceases and amplification is thus limited to a magnetospheric volume characterized by the three scale lengths L_r in radius, $L_\perp$ in azimuth (both in the plane perpendicular to $\mathbf{B}_0$) and L_z parallel to the magnetic field. These parameters are now estimated for a dipole magnetic field.

The radial dimension L_r derived in ref. 3 is given below with the scale length L_{rp} due to plasma frequency variations assumed inversely proportional to R^2 (ref. 11)

$$L_r \approx \frac{R_0 \Delta\bar{\omega}}{3 \bar{\omega}} \quad L_{rp} \approx \frac{R_0 \Delta\omega'}{2 \omega'}$$

where $\omega' = \omega/\omega_p$. In the model $\Delta\bar{\omega}/\bar{\omega} \approx \Delta\omega'/\omega'$ and thus for the same percentage variation in $\bar{\omega}$ and ω' magnetic field variations yield the smallest scale length. Moreover calculations for different plasma densities show that the dispersion curves in the region corresponding to large temporal growth rates, $k_\perp R \approx 30$, are less sensitive to plasma density variations than the upper hybrid frequency is. Therefore magnetic field variations make L_r the dominant quantity. The azimuthal dimension $L_\perp$ is obtained by adopting the ray tracing method of Barbosa and Kurth¹² where the spatial variation of wave refractive index $n(\bar{\omega})$ due to magnetic field changes is expressed as a power law in $\bar{\omega}$.

$$n(\bar{\omega}) = \bar{\omega}^\alpha$$

Here α is the power index evaluated over the perpendicular wavenumber range corresponding to instability. The result is

$$L_\perp \approx R_0 \left(\frac{8\Delta\bar{\omega}}{3\bar{\omega} (3|\alpha| \pm 1)} \right)^{\frac{1}{2}}$$

where $R_0 = 6.6$ Earth radii and the negative sign should be taken for $\alpha > 0$. In the calculations $\Delta\bar{\omega}$ is identified with the frequency bandwidth over which significant spatial amplification exists, with the results $L_\perp \approx 600$ km for model A and $L_\perp \approx 1,500$ km for model B. The parallel dimension, L_z , for this particular class of wave activity is taken from observations. Strongest signals have been reported to be confined to $\pm 2.5^\circ$ about the geomagnetic equator⁴. More recent observations¹³ show the confinement to be as small as $\pm 1^\circ$ from which the value $L_z = 1,500$ km is used in these calculations. Note that equatorial confinement of the upper hybrid emission is probably due to a combination of refraction effects and loss of resonance in weak field gradients¹².

Computations of the group velocities show that the waves traverse the amplifying region in a few tens of seconds at all frequencies except $\sim 6.6 f_{ce}$. At this frequency, which corresponds to the maximum in the dispersion curves, the perpendicular group velocity changes sign while the parallel group velocity remains small. Predictions for model B then indicate that such a wave, which is essentially absolutely unstable, would remain in the amplifying region for some tens of minutes. That

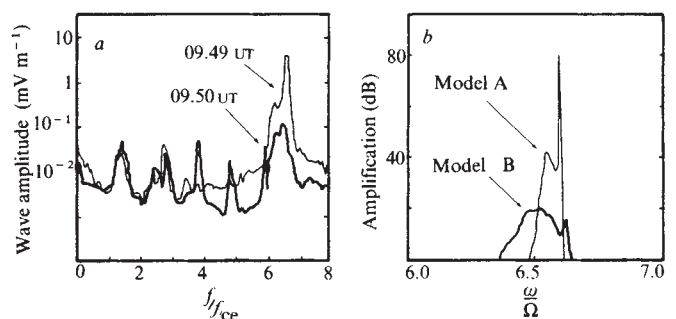


Fig. 3 Comparison between the observed wave spectra at 09.49 and 09.50 UT (a) and the calculated wave amplification between 6Ω and 7Ω (b). Strong electrostatic waves are not observed between $7f_{ce}$ and $8f_{ce}$ and the cold upper hybrid frequency is estimated at $6.05 f_{ce}$ (see text).

this cannot happen can be seen from the observation that the duration of both strong and weak emissions is ~ 90 s. After this period changes in the particle distribution remove the near absolute instability and increase both parallel and perpendicular group velocities.

Furthermore, detuning effects arising from wave propagation in the z direction, which are not considered in the ray tracing model above, will also occur.

The final amplification levels predicted for both models are shown in Fig. 3. Convective amplification levels have been calculated in the same way as in ref. 3 using the scale lengths deduced above. For frequencies corresponding to near absolute instability the temporal amplification is assumed to be limited to ~ 90 s, irrespective of whether the waves have traversed the amplifying region or not. Figure 3 also shows wave spectra representative of the two emission levels.

The prediction for model B is a spectrum peaking at the 20-dB amplification level, while that of model A exceeds it by at least 20 dB and rises to a maximum of 80 dB amplification. A significant difference is also seen in the natural wave spectrum where the maximum amplitude levels in the $6-7f_{ce}$ band for the strong emissions (corresponding to model A) exceed that of the model B related intermediate level emissions by at least 30 dB.

We conclude that the model used in this study includes the main features necessary to explain the sporadic transitions between the intermediate ($E > 100 \mu\text{V m}^{-1}$) and large amplitudes ($E > 3 \text{ mV m}^{-1}$) typical of dayside ECH emissions. These transitions are due to subtle changes in the thermal and suprathermal electron populations, and this emphasizes that detailed understanding of such emissions requires accurate information about the distribution of the suprathermal population as well as that of the 'free-energy' source. Furthermore, even if there are small changes in the free energy source, the prediction of such a large difference between the maximum model amplification levels suggests that this conclusion remains unchanged. Further calculations designed to test the sensitivity of the models to small variations in the value of the cold plasma density support these conclusions. Predictions for ω_{uh} as low as 5.8Ω show that the difference between the maximum amplification levels in the $6-7\Omega$ band is maintained, whereas for $\omega_{uh} > 6.6\Omega$ there is virtually no difference between the levels in the $6-7\Omega$ band but large amplitudes are also predicted above 7Ω . As strong emissions above 7Ω are not observed experimentally, we conclude that ω_{uh} lies close to 6Ω and that the value used here of $\omega_{uh} = 6.05\Omega$ is realistic.

Finally, we point out that many observations of strong sporadic emissions have been made using the GEOS 2 satellite, and the appearance of suprathermal, approximately field-aligned electron components, presumably of ionospheric origin, often show strong anticorrelation with the upper hybrid band emission amplitudes.

We thank Mrs E. J. Gershuny for help in data handling and Dr G. Martelli for his interest. This research is supported by SRC, the Swedish Board for Space Activities and the ESA.

Crystal structure of a synthetic high silica zeolite—ZSM-39

J. L. Schlenker, F. G. Dwyer, E. E. Jenkins,
W. J. Rohrbach, & G. T. Kokotailo

Mobil Research and Development Corporation,
Research Department, Paulsboro, New Jersey 08066 USA

W. M. Meier

Institute of Crystallography and Petrography, ETH,
Sonneggstrasse 8, 8092 Zurich, Switzerland

The crystal structure of a high silica zeolite ZSM-39, unit cell composition excluding residual water and occluded materials— $(\text{Na}, \text{tetramethylammonium ion}, \text{tetraethylammonium ion})_{0.4}(\text{AlO}_2)_{0.4}(\text{SiO}_2)_{135.6}$, was determined by X-ray powder diffraction. The framework is pseudoface-centred, pseudocubic with $a = 19.36 \pm 0.02 \text{ \AA}$ and ideal symmetry $\text{Fd}3\text{m}$. The framework consists of a space-filling arrangement of pentagonal dodecahedra and hexakaidecahedra and is isostructural with the $17 \text{ \AA}$ cubic gas hydrate. The ZSM-39 framework is composed entirely of five- and six-rings which limits its sorptive and exchange properties. However, the large fraction of five-rings and the high Si/Al ratio (>40) impart a high thermal stability. ZSM-39 containing no aluminium constitutes the end member composition. Although ZSM-39 is the only synthetic zeolite analogue of a gas hydrate, we propose here two related hypothetical frameworks containing pentagonal dodecahedral cages.

The analogy between hydrogen bonds linking oxygens in clathrates and oxygens linking T-atoms in aluminosilicate framework structures is well known. Hexagonal ice I and ice Ic are isostructural with β -tridymite and β -cristobalite respectively¹, $\text{HPF}_6 \cdot 6\text{H}_2\text{O}$ and $(\text{CH}_3)_4\text{N}(\text{OH}) \cdot 5\text{H}_2\text{O}$ are isostructural with sodalite^{2,3} and the rare mineral, melanophlogite⁴⁻⁶ is isostructural with the $12 \text{ \AA}$ cubic gas hydrate^{7,8}. ZSM-39 is the first zeolite which is isostructural with the $17 \text{ \AA}$ cubic gas hydrate^{9,10}.

ZSM-39 was synthesized hydrothermally in a system containing tetramethylammonium and tetraethylammonium hydroxide, silica, and sodium aluminate. Crystallization was at $\sim 155^\circ\text{C}$ and was complete after 330 h (ref. 11). Analysis of a ZSM-39 sample gave (wt% composition): SiO_2 , 89.0%; Al_2O_3 , 00.2%; Na_2O , 00.3%; N, 00.8%; C, 03.76%. The remaining 5% could not be accounted for with complete confidence. The analytical data refer to the as-synthesized material before calcination. In this state the unspecified difference can be attributed primarily to water and hydroxyl groups.

Crystals of ZSM-39 tend to have octahedral habit as shown in Fig. 1. Single crystals $>10 \mu\text{m}$ could not be grown. ZSM-39 has

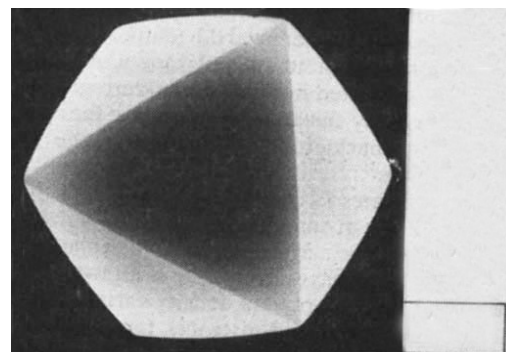


Fig. 1 Scanning electron micrograph of ZSM-39 single crystal. The habit suggests cubic symmetry. $\times 4,500$.

Received 14 November 1980; accepted 26 August 1980.

- Kennel, C. F., Scarf, F. L., Fredricks, R. W., McGehee, J. H. & Coroniti, F. V. *J. geophys. Res.* **75**, 6136 (1970).
- Lyons, R. L. *J. Geophys. Res.* **79**, 575 (1974).
- Rönmark, K. G. & Christiansen, P. J. *Nature* **294**, 335–338 (1981).
- Gough, M. P., Christiansen, Martelli, F. & Gershuny, E. J. *Nature* **279**, 515 (1979).
- Christiansen, P. J. *et al. Space Sci. Rev.* **22**, 383 (1978).
- Johnson, J. F. E., Sojka, J. J. & Wrenn, G. L. *Space Sci. Rev.* **22**, 576 (1978).
- Borg, H. *et al. Space Sci. Rev.* **22**, 511 (1978).
- Horne, R. B. *et al. Advances in Space Research* Vol. 1 (ed. Knott, K.) (Pergamon, Oxford, 1981).
- Rönmark, K. G., Borg, H., Christiansen, P. J., Gough, M. P. & Jones, D. *Space Sci. Rev.* **22**, 401 (1978).
- Ashour-Abdalla, M. & Kennel, C. F. in *Magnetospheric Particles and Fields* (ed. McCormac, B.) 181 (Riedel, Dordrecht, 1976).
- Kurth, W. S., Craven, J. D., Frank, L. A. & Gurnett, D. A. *J. geophys. Res.* **84**, 4145 (1979).
- Barbosa, D. D. & Kurth, W. S. *J. geophys. Res.* **85**, 6729 (1980).
- Gough, M. P., Christiansen, P. J. & Gershuny, E. J. in *Advances in Space Research* Vol. 1 (ed. Knott, K.) (Pergamon, Oxford, 1981).

Table 1 X-ray diffraction data for ZSM-39

HKL	2 θ (deg)	d_{obs} (Å)	d_{calc} (Å)	Relative intensity
None	7.50	11.79	—	1
111	7.93	11.15	11.18	5
220	12.95	6.84	6.84	23
None	13.72	6.45	—	2
None	14.35	6.17	—	1
311	15.21	5.83	5.84	93
222	15.88	5.58	5.59	69
None	16.58	5.35	—	1
400	18.37	4.83	4.84	47
331	20.03	4.43	4.44	36
None	21.20	4.19	—	2
422	22.51	3.95	3.95	48
331; 511	23.90	3.72	3.73	100
None	24.55	3.63	—	1
440	26.04	3.42	3.42	42
531	27.25	3.27	3.27	84
442; 600	27.63	3.23	3.23	10
620	29.16	3.06	3.06	12
533	30.26	2.95	2.95	8
551; 711	33.05	2.71	2.71	2
553; 731	35.63	2.52	2.52	9
800	37.15	2.42	2.42	1
733	38.05	2.36	2.36	10
660; 822	39.51	2.28	2.28	17
555; 751	40.34	2.24	2.24	3
662	40.64	2.22	2.22	1
840	41.70	2.17	2.16	3
753; 911	42.53	2.13	2.12	1
844	45.96	1.98	1.98	3
771; 755; 933	46.67	1.95	1.95	4
862; 10, 20	47.91	1.90	1.90	2
666; 10, 22	48.88	1.86	1.86	10
953	50.55	1.81	1.80	6
10, 42	51.72	1.77	1.77	5
955; 971; 11, 31	54.24	1.69	1.69	3
882; 10, 44	54.47	1.68	1.68	4
10, 60	55.35	1.66	1.66	5
973; 11, 31	56.02	1.64	1.64	2
12, 00; 884	57.10	1.61	1.61	5
11, 51; 777	57.72	1.60	1.60	8
10, 64; 12, 22	58.80	1.57	1.57	10
975; 11, 53	59.44	1.56	1.56	5

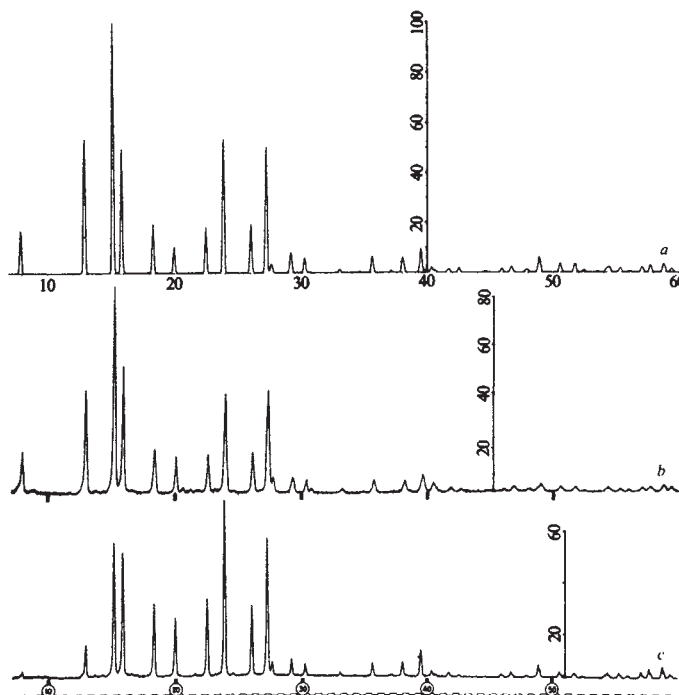
Table 2 Atomic coordinates of ZSM-39 Fd3m with origin at $\bar{3}m$

	x	y	z
T1	0.125	0.125	0.125
T2	0.2195	0.2195	0.2195
T3	0.1823	0.1823	0.3728
O1	0.1722	0.1722	0.1722
O2	0.2024	0.2024	0.2967
O3	0.125	0.125	0.3743
O4	0.25	0.1573	0.4073

little zeolitic sorptive capacity even after calcination at high temperatures. The material becomes black on calcination indicating that carbonaceous material is being retained in the pore structure. The same effect is observed with melanophlogite¹. The thermal stability is very high and crystallinity is retained even after calcination at temperatures above 800 °C.

X-ray diffraction patterns of ZSM-39 were obtained using a Norelco powder diffractometer and CuK α radiation. Powder diffraction data are given in Table 1.

Morphology, X-ray powder, and sorption data were used to determine the structure. Morphological evidence indicative of cubic symmetry was confirmed by analysis of the powder pattern and a cell parameter of 19.36 ± 0.02 Å (D. H. Olson, personal communication) was obtained by least-squares refinement using 39 indexed reflections. Except for a few very weak lines, all reflections attributable to ZSM-39 are consistent with Fd3m symmetry. The high symmetry and lack of sorptive capacity suggested a clathrate-like structure. Examination of these structures revealed that the 17 Å cubic gas hydrate⁹ had ideal symmetry Fd3m. This structure has an average lattice constant of 17.2 Å and a hydrogen bond distance of 2.73 Å. Assuming an

**Fig. 2** a, Simulated powder pattern of the ZSM-39 framework. b, Powder pattern of ZSM-39 (sample after calcination for 0.5 h at 800 °C). c, Powder pattern of ZSM-39 (sample as synthesized).

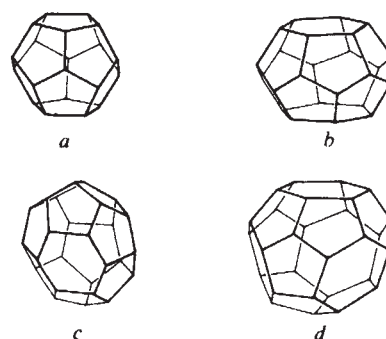
Si—O—Si bond distance of 3.07 Å, the postulated adjusted cell parameter of the silica analogue would be 19.34 Å which compares well with the observed value of 19.36 Å for ZSM-39.

Least-squares atomic positions for the T and oxygen atoms, obtained using the DLS method¹² are given in Table 2. A Smith plot¹³ obtained using these coordinates is compared with an X-ray powder pattern of ZSM-39 in Fig. 2. The agreement indicates that the ZSM-39 framework is topologically equivalent to that of the 17 Å cubic gas hydrate.

A number of very weak reflections inconsistent with a face-centred lattice indicated that the symmetry of ZSM-39 is lower than Fd3m. This is similar to the situation with melanophlogite¹⁴.

The framework of ZSM-39 consists of 12-hedra (pentagonal dodecahedra) and 16-hedra (hexakaidecahedra) as shown in Fig. 3. The structure is composed of layers of face-sharing 12-hedra arranged as in Fig. 4a. In ZSM-39 these layers are stacked in an ABC sequence. The framework may be alternately viewed as 16-hedra linked tetrahedrally through common six-rings. The arrangement of 16-hedra is the same as that of the carbon atoms in diamond.

The layers of ZSM-39 (Fig. 4a), may also be stacked in an AB sequence, giving rise to a hypothetical framework with ideal

**Fig. 3** Polyhedral building units of the ZSM-39 and melanophlogite structures. a, 12-hedron (dodecahedron); b, 14-hedron (tetrakaidecahedron); c, 15-hedron (pentakaidecahedron); d, 16-hedron (hexakaidecahedron).

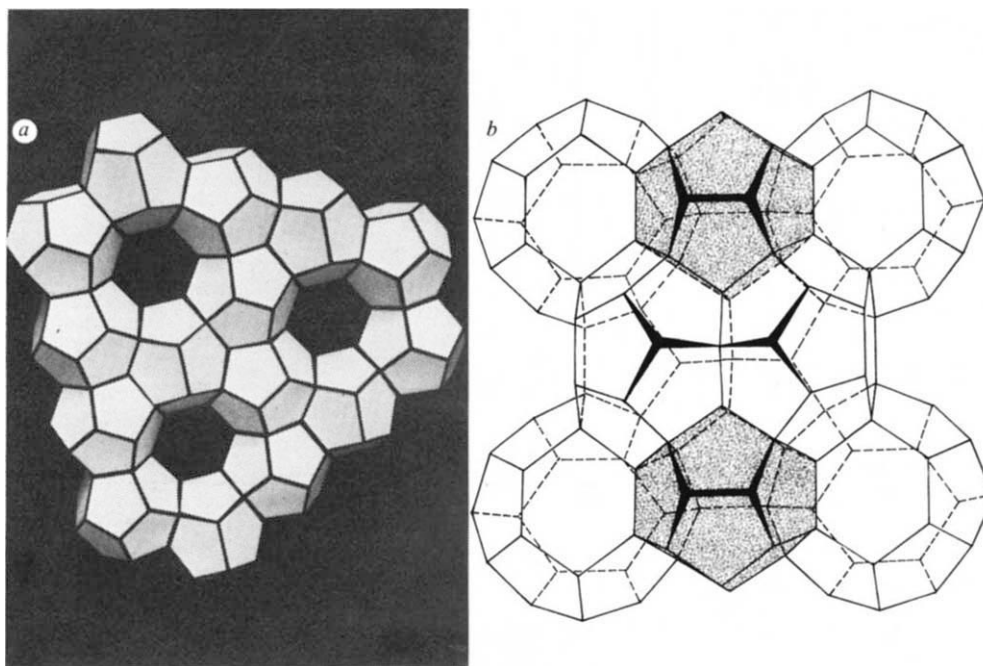


Fig. 4 *a*, Layer of ZSM-39 framework illustrating arrangement of face-sharing 12-hedra. *b*, Melanophlogite framework illustrating isolated 12-hedra.

symmetry, $P\bar{6}m2$, and cell parameters of $a = 12.2$ and $c = 24.2$ Å. This structure contains the same polyhedra as ZSM-39. These two frameworks are the simplest members of a polytypic series.

Melanophlogite, the first reported silica analogue of a gas hydrate, has a framework consisting of interwoven layers of 12- and 14-hedra (Fig. 3). In contrast to ZSM-39, this framework contains isolated 12-hedra (Fig. 4*b*). Sliding alternate layers a distance of $a/2$ results in a hexagonal framework with ideal symmetry $P6/mmm$ and $a = 13.6$ and $c = 14.0$ Å. This hypothetical framework consists of a space-filling arrangement of 12-, 14-, and 15-hedra (Fig. 3). Although there is no known zeolite of this type, $(i-C_5H_{11})_4NF \cdot 38H_2O$ does adopt this structure³.

Porotectosilicate analogues of water-clathrate structures which typically consist of cages comprising pentagonal rings constitute a new family of high-silica materials, most of which are still hypothetical. ZSM-39, the first synthetic member, should encourage further attempts to synthesize other members of the family including melanophlogite, a small molecule trap that has long intrigued mineralogists.

We thank D. H. Olson, S. L. Lawton, and P. Chu for assistance, J. B. Milliken and A. C. Rohrman Jr for density data and SEM micrographs, also G. A. Jeffrey and D. E. Appleman for their valuable discussions regarding clathrates and melanophlogite, respectively. W.M.M. acknowledges the support of the Swiss NSF.

Received 13 July; accepted 2 October 1981.

1. Kamb, B. *Science* **148**, 232 (1965).
2. McMullon, R. K., Mak, T. C. W. & Jeffrey, G. A. *J. chem. Phys.* **44**, 2338 (1966).
3. Feil, D. & Jeffrey, G. A. *J. chem. Phys.* **35**, 1863 (1961).
4. von Lasaux, A. *Neues Jb. Miner. Abh.* **250**, (1876).
5. Skinner, B. J. & Appleman, D. E. *Am. Miner.* **218**, 854 (1963).
6. Appleman, D. E., *Amer. Crystallogr. Ass., Mineral. Soc. Amer. Jt Meet., Gatlinburg*, 80 (1965).
7. Pauling, L. & Marsh, R. E., *Proc. natn. Acad. Sci. U.S.A.* **38**, 112 (1952).
8. Claussen, W. F. *J. chem. Phys.* **19**, 1425 (1951).
9. Claussen, W. F. *J. chem. Phys.* **19**, 112 (1951).
10. Gros, G., Ponchard, M., Hagenmueller, P. *J. Solid-State Chem.* **2**, 570 (1970).
11. Dwyer, F. G. & Jenkins, E. E. U.S. Patent 4,287,166 (1981).
12. Baerlocher, Ch., Hepp, A. & Meier, W. M. *DLS-76, A Program for the Simulation of Crystal Structures by Geometric Refinement* (Zurich, 1977).
13. Smith, D. K. *A Revised Program for Calculating X-ray Powder Diffraction Patterns UCRL-502.64* (Lawrence Radiation Laboratories, 1967).
14. Zák, L. *Am. Miner.* **57**, 779 (1972).

Hydrothermal leaching of rhyolite glass in the environment has implications for nuclear waste disposal

A. P. Dickin*

Department of Geology and Mineralogy, University of Oxford, Oxford OX1 3PR, UK

Glass has been widely advocated as a suitable medium for the immobilization of high-level nuclear waste¹⁻⁶. Methods of vitrification to borosilicate glass are advanced, with processes set up on a semi-industrial scale⁷⁻⁹, but an alternative strategy would be to incorporate waste into a high-silica glass¹⁰. It is generally proposed^{5,6} that vitrified radioactive waste be enclosed in metal canisters and stored underground. However, the presence of heated groundwater in rocks means that the canisters may corrode, allowing hot aqueous fluids to come into contact with the radioactive waste glass and leach out radioactive elements. Numerous laboratory tests have been performed for short periods and at relatively low temperatures to assess the leaching performances of different types of glass¹¹⁻¹⁵, however, extrapolation to predict the long-term behaviour of glasses after burial is very uncertain. I report here a microprobe analysis technique which investigated hydrothermal leaching of rhyolite glass adjacent to a fluid conduit in the Tertiary hydrothermal system of the Isle of Skye, north-west Scotland. As the composition of rhyolite and proposed high-silica radioactive waste glasses are similar, this study may help to predict the long-term leaching behaviour of such glasses after underground burial.

Fractures which formed fluid conduits for the Skye hydrothermal system have been plugged by secondary mineral growth, and now form veins which cut through the rocks. The analysed sample is cut by such a hydrothermal vein, and comes from near the centre of an acid/basic composite sill. This was exposed in 1979 in a roadstone quarry at Sconser (NG 544 318) on the Isle of Skye (Fig. 1*b*). The sill is intruded into the late Precambrian 'Torridonian' arkosic sandstones only 300 m from the northern margin of the Western Redhills centre of granitic

* Present address: Scottish Universities Research and Reactor Centre, East Kilbride, Glasgow G75 0QU, UK.

intrusion¹⁶. A gravity survey¹⁷ (Fig. 1c) suggests that the margin of the granite complex is almost vertical, and passes directly downwards into the margin of a massive basic pluton underlying Skye. The heat evolved from this body drove a huge meteoric-hydrothermal convection system, whose existence is demonstrated by very depleted $\delta^{18}\text{O}$ compositions in the rocks in and around the central complex¹⁸. These $\delta^{18}\text{O}$ compositions closely resemble those in the vicinity of the Skaergaard intrusion of Western Greenland¹⁹, suggesting that numerical modelling of the Skaergaard hydrothermal system²⁰ is applicable to Skye.

Epidote-calcite veins which cut Torridonian sandstone in Sconser quarry continue through the Tertiary sill without interruption, indicating that a single hydrothermal event affected both sill and country rocks. Therefore, the sill must have been subjected to the full duration of the hydrothermal convection system generated by the plutonic complex. Veins are predominantly aligned radially to the Redhills complex, and those which are similar in size to the studied example have a separation of ~ 10 cm (ref. 21). Assuming a separation of 10 cm between parallel planar fractures, comparison with Norton and Taylor's S2 numerical model of the Skaergaard suggests an integrated fluid flux of over 10,000 l per cm-fracture-length, almost all of which occurred in an initial 200,000 yr period at $\sim 400^\circ\text{C}$. An integrated flux of this magnitude provides an effectively open system for alteration of the fracture wall rocks, ruling out the possibility that fluid composition could be buffered within the sill.

The presence of tridymite in the Coire Uaighneigh Granophyre of Skye led Brown²² to deduce a value of 1.5 km for maximum post-Tertiary erosion of the Skye lava pile. As Scon-

ser quarry lies near sea level, this suggests that the hydrothermal vein under investigation was at a depth > 1 km, corresponding to a lithostatic pressure of 0.3 kbar or a hydrostatic pressure of 0.1 kbar. The actual pressure in the vein was probably between these two limits.

A low-power photomicrograph of the vein (Fig. 2a) shows it to be a nearly planar, sharply bounded fracture, 0.9 mm wide, filled by large sparry calcite crystals and fringed by epidote. This vein must once have been an open fissure, as euhedral prismatic epidote crystals grow out from the walls into the free space of the fracture (Fig. 2b).

Figure 2c shows wall rock material a few mm from the vein under high power and crossed polars. Aggregates of small alkali feldspar and quartz crystals are interspersed by patches of a more isotropic matrix, showing that the microcrystalline felsitic texture of the rock was produced by devitrification of an original glass. The grain size of the felsite falls slightly in a zone of bleached rock within 1.5 mm of the fracture. Devitrification of the original glass probably occurred during the period of hydrothermal metamorphism, rather than before or after. In this period, the presence of hot fluids would have speeded up kinetic processes of recrystallization. Figure 2d shows a map of K distribution in the felsite which picks out the distribution of alkali feldspar.

A polished section of typical wall-rock material, perpendicular to the plane of the fracture, was analysed using an electron microprobe (Cambridge Scientific Instruments Microscan 9). The probe beam of $6 \times 10^{-8}\text{A}$ at 20 kV was rastered over $100 \mu\text{m} \times 100 \mu\text{m}$ squares, which were analysed in 2.5 mm long strips, stretching from the vein wall into unbleached material. Twelve adjacent strips were analysed, and mean compositions were calculated for each increment of distance away from the vein. For each oxide, results outside two standard deviations from the mean were rejected, and if three or more oxides were thus excluded then the whole analysis was discounted. This procedure was adopted in an attempt to remove the effects of microphenocrysts from the felsite analyses. The area analysed was chosen for the absence of adjacent large phenocrysts.

Modal mineral analyses were determined by 'point counting' at $40 \mu\text{m}$ intervals, using 2,500 partial microprobe analyses. The beam was focused to spot, resulting in an analysis area of a few square micrometres. The zone thus analysed was the same as for the whole-rocks.

Whole-rock microprobe analyses are plotted in Fig. 3 as profiles of oxide weight percentages against distance from the vein wall. As the vein is approached, SiO_2 and MgO contents plunge, while Al_2O_3 , CaO and Na_2O rise. These compositional variations extend up to a distance of ~ 1.5 mm from the vein wall, beyond which major-element variations are not correlated with distance. Contents of K_2O , MnO , P_2O_5 , SrO and BaO remain essentially constant from 0 to 2.5 mm from the vein, while total iron (expressed as FeO) and TiO_2 vary rather erratically, due to the uneven distribution of titanomagnetite microphenocrysts (Fig. 2a).

To quantify the oxide variations as a function of distance from the vein, weight percentages have been regressed against distance, and correlation coefficients calculated. The results are shown in Table 1. Contents of SiO_2 and MgO display very strong positive correlations with distance from the vein, while Al_2O_3 , CaO and Na_2O display very strong negative correlations with distance. SiO_2 , Al_2O_3 , CaO and Na_2O form essentially linear trends between 0 and 1.5 mm from the vein, while MgO drops dramatically at 1.0 mm, before bottoming out within 0.5 mm of the vein. FeO and deficit-of-totals-below-100% ($=\text{H}_2\text{O} + \text{CO}_2$) correlate more weakly with distance from the vein, while the correlation of TiO_2 , K_2O , SrO and BaO with distance is highly insignificant. The theoretical composition of the vein wall is calculated from the intercept of the regression lines at 0.0 mm from the vein, while the mean composition of felsite between 1.8 and 2.5 mm has also been calculated. The latter compares quite closely with the composition of 'world average rhyolite' (Table 1)²³. The silica content at the vein wall (0.0 mm) is 9.34 wt% less

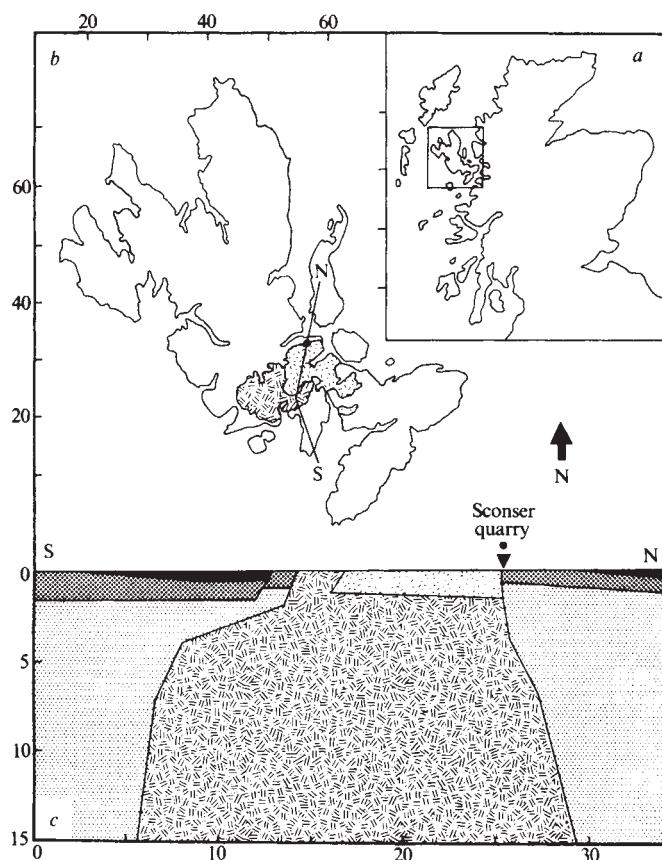


Fig. 1 a, Map of Scotland showing the location of the Isle of Skye. b, Map of Skye showing the Cuillin basic pluton (double hatching) and Redhills granites (single hatching). The spot marks the location of Sconser quarry, which lies on the line of section marked N/S. Ticks mark 10-km national grid squares. c, Crustal section through Skye based on geophysical data¹⁷. Ornament for intrusive rocks as in b. Country rocks: light stipple, Lewisian (Archaean) gneiss; heavy stipple, Torridonian sandstone; black, Mesozoic sediments. Horizontal and vertical distances in km.

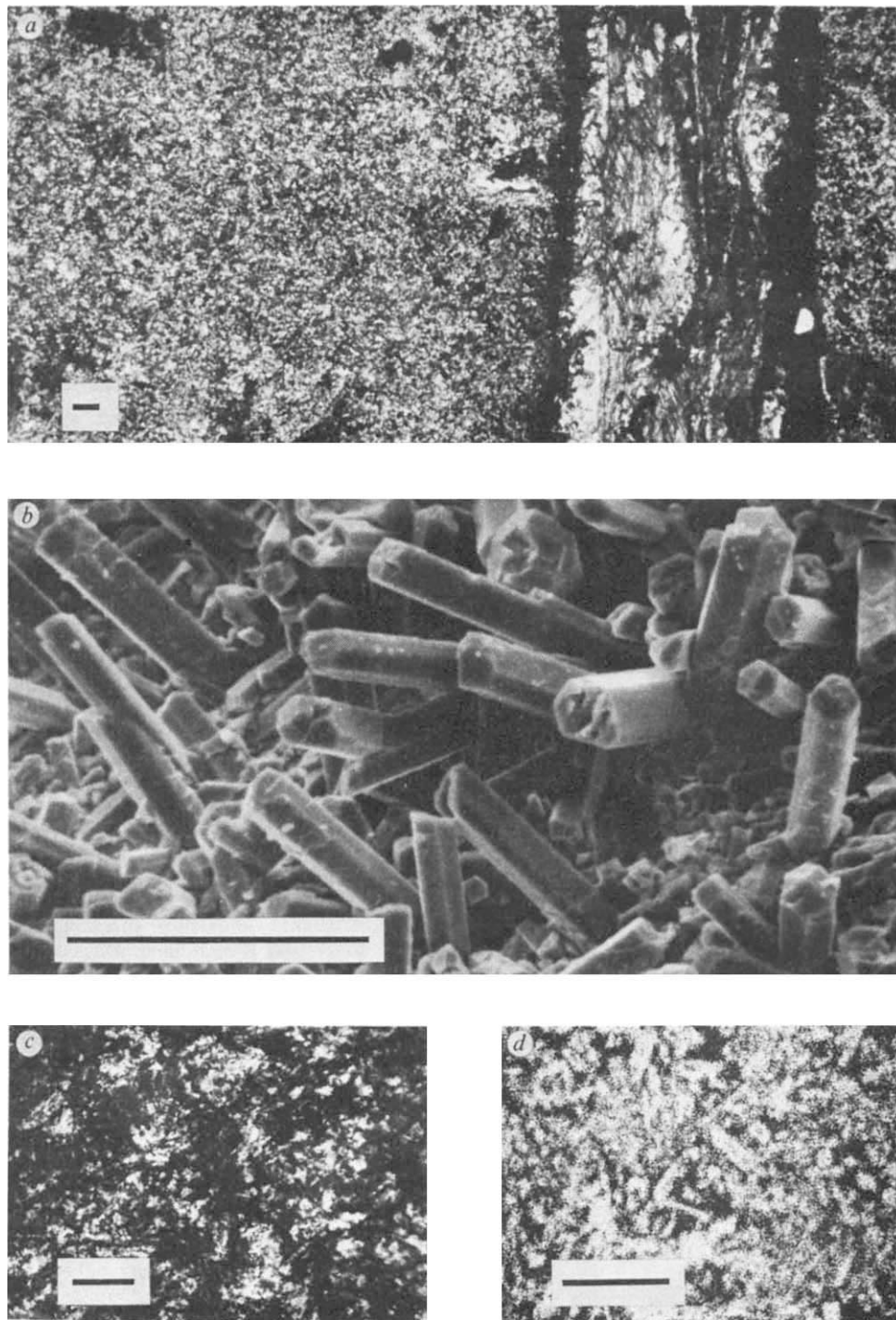


Fig. 2 *a*, Low-power photomicrograph of vein and analysed wall rock in plane polarized light. *b*, Scanning electron micrograph of prismatic epidotes growing from vein wall into fracture, after removal of vein filling calcite by dilute HCl. *c*, High-power view of feldspar a few millimetres from vein showing devitrification texture, under crossed polars. *d*, X-ray backscatter picture showing K distribution in feldspar 2 mm from vein. Scale bars, 100 μm .

than in the unaltered material (1.8–2.5 mm), while the maximum enrichments of Al_2O_3 , CaO and Na_2O at the vein wall sum to 9.14%. However, the build-up of these oxides cannot be a relative concentration effect caused simply by silica removal, because the enrichment factors (as a proportion of the fresh material) are different for each oxide (Table 1).

Whole-rock compositional variations in the vicinity of the vein are ascribed to exchange of species between the wall rocks and meteoric-hydrothermal fluids passing up the vein. Nevertheless, it is not immediately clear whether the changes are a product of fluid composition, or of mineralogical constraints. To assess the effects of mineralogy on whole-rock composition, modal mineral determinations were made by microprobe 'point counting'. The results are shown in Fig. 4. The abundances of calcite, epidote, quartz and alkali feldspar display correlations with distance in the region 0–1.5 mm from the vein, while beyond this distance the profiles are essentially flat. These

results are consistent with the whole-rock analyses and have similarly been regressed against distance from the vein (see Table 1).

Modal alkali feldspar content rises from an average value of 74.3% (1.8–2.5 mm from vein) to a value of 90.7% at the vein wall. Similarly, epidote increases from 0.8 to 7.5% and calcite rises from 0.5 to 1.1%. In contrast, the modal quartz content falls from 22.1% to a calculated value of -0.7% (± 1.8) at the vein wall. All these elements correlate very strongly with distance from the vein ($>99\%$ confidence), but modal oxide contents do not correlate with distance at all ($r^2 = 0.21$). Normative mineralogies calculated from the whole-rock compositions are consistent with modal mineral abundances. However, separation of alkali feldspar into its orthoclase and albite components shows orthoclase to be constant (31%) with distance, while albite rises sharply from 35 to 57% as the vein is approached.

Table 1 Oxide and mineral composition variations as a function of distance from the vein

	World average rhyolite	Average felsite 1.8–2.5 mm	Intercept at vein wall, 0 mm	Intercept minus 1.8–2.5 mm	Enrichment factor at vein wall	Correlation coefficient	No. of data points regressed
Oxide							
SiO ₂	72.82	72.59	63.25	−9.34	−13	+0.98	15
Al ₂ O ₃	13.27	13.50	18.45	+4.95	+37	−0.98	15
TiO ₂	0.28	0.09	0.10	+0.01	—	−0.04	21
Fe ₂ O ₃	1.48	—	—	—	—	—	—
FeO	1.11	1.18	0.70	−0.48	−41	+0.66	21
MnO	0.06	0.03	0.03	0	—	ND	—
MgO	0.39	0.14	0.02	−0.12	−86	+0.98	6*
CaO	1.14	0.50	2.03	+1.53	+306	−0.84	15
Na ₂ O	3.55	4.12	6.78	+2.66	+65	−0.98	15
K ₂ O	4.30	5.25	5.38	+0.13	—	−0.16	21
P ₂ O ₅	0.07	0.01	0.01	0	—	ND	—
SrO	ND	0.022	0.027	+0.005	—	−0.20	15
BaO	ND	0.125	0.141	+0.016	—	−0.22	24
H ₂ O+CO ₂	1.49	2.44	2.97	+0.53	+22	−0.64	20
Total	99.96	100	99.89	—	—	—	—
Mineral							
Calcite		0.5	1.1	+0.6	+120	−0.71	14
Oxide		2.2	1.0	−1.2	—	+0.21	14
Epidote		0.8	7.5	+6.7	+838	−0.79	14
Quartz		22.1	−0.7	−22.8	−103	+0.92	14
Alkali feldspar		74.3	90.7	+16.4	+22	−0.79	14

ND, not determined.

* Not starting at vein wall.

Between 1.8 and 2.5 mm from the vein, devitrification of glass was isochemical, and formed a typical 'granitic' mineralogical assemblage of alkali feldspar and quartz with a subordinate hydrous ferromagnesian component. However, in the conditions of high water/rock ratio found at the vein wall, albite growth was particularly favoured. This developed by devitrification of glass, but also 'scavenged' additional Al₂O₃ and Na₂O from hydrothermal fluids. In contrast, silica in the glass which was in excess of the requirements of alkali feldspar was removed by hydrothermal fluids as fast as devitrification occurred, and was not able to form quartz. The development of K feldspar seems to have been independent of water/rock ratio, as it is effectively the same at the vein wall as between 1.8 and 2.5 mm from the vein. This may also explain the flatness of the BaO profile, as the Ba²⁺ ion is too big to fit into albite but is readily incorporated into orthoclase²⁴. Epidote growth was strongly favoured at the vein wall, and incorporated iron, SiO₂, Al₂O₃ and CaO from the glass, but 'scavenged' additional CaO from hydrothermal fluids. The abundance of CaO in these fluids is indicated by the plugging of the vein with calcite, which does not form a major component in the altered wall rock.

McCarthy *et al.*²⁵ noted that most leachability tests on borosilicate glass are conducted at 25 °C and 1 atm, whereas temperatures²⁶ in underground waste repositories could reach 400 °C. In view of these dangers, the proposed radioactive waste content of glasses has been reduced from its original 25% (ref. 11), but the cores of full size 'production' waste blocks will almost certainly exceed temperatures of 100 °C for long periods of time. In these conditions it cannot be assumed that the metal canister enclosing a waste block will not be breached by corrosion. Observed fracture development in such blocks yields glass lumps of average radius 5 cm (ref. 27), and ensures that if such a breach does occur then fluids can penetrate to the high-temperature core of the waste block.

A leaching experiment on 11-yr old 'Fingal' radioactive waste glass at 90 °C and 1 bar revealed 0.0023 g cm^{−2} of leaching in 1 week, yielding an estimated linear leaching rate of 40 cm in 1,000 yr (ref. 11). This would be sufficient to disintegrate a waste block fractured into 5-cm radius pieces in <100 yr. Similarly, McCarthy *et al.*²⁵ found that 4-mm spheroids of simulated-waste borosilicate glass broke into fragments and 'seemed to be totally

altered' after only 2 weeks at ~300 °C and 300 bar. In contrast, integration of the degree of oxide depletion in the rock analysed in this work, between the vein wall and a distance of 1.5 mm away, indicates that the total quantity of material leached by fluids per cm² of vein wall was 0.019 g. This occurred in the presence of a high fluid flux at 400 °C and 300 bar, which lasted perhaps 200,000 yr. These comparisons show that high-silica glass has a much higher resistance to high temperature hydrothermal leaching than borosilicate glass.

Karkhanis *et al.*¹⁵ found that disks of rhyolite glass leached with heated distilled water produced large amounts of muscovite, whereas those reacted in the presence of granite showed only a small amount of crystalline product which appeared to be feldspar. This observation is consistent with the present model of alteration processes; formation of the stable alkali-feldspar/epidote alteration assemblage being made possible by 'scavenging' of Ca²⁺, Na⁺ and Al³⁺ from hydrothermal fluids. At Sconser, the hydrothermal fluids must have picked up these species by interaction with the arkosic Torridian sandstone country rocks. These sediments have compositions close to the granite eutectic²⁸, suggesting that granitic rocks would form a favourable geochemical environment for burial of high-silica radioactive waste glass.

It is important to assess the effect that addition of radioactive elements would have had on the stability and leach-resistance of the glass. This would result in significant radiation damage to the structure of glasses, and tend to precipitate devitrification. However, the rock analysed in this work is an original glass which has suffered more or less complete devitrification in a high temperature aqueous environment. This more than equals the degree of devitrification that might be induced by radiation damage.

It is also necessary to consider whether leaching of radioactive elements would be significantly greater than the analysed major elements. The radioactive elements fall into three groups: low atomic number fission products, high atomic number fission products, and actinides. Isotopic investigations of the mobility of Sr (a low atomic number fission product), during hydrothermal alteration of a Skye granophyre²⁹, showed that Sr mobility was only significant in the immediate vicinity of hydrothermal veins, because it was strongly partitioned into alkali feldspar. The

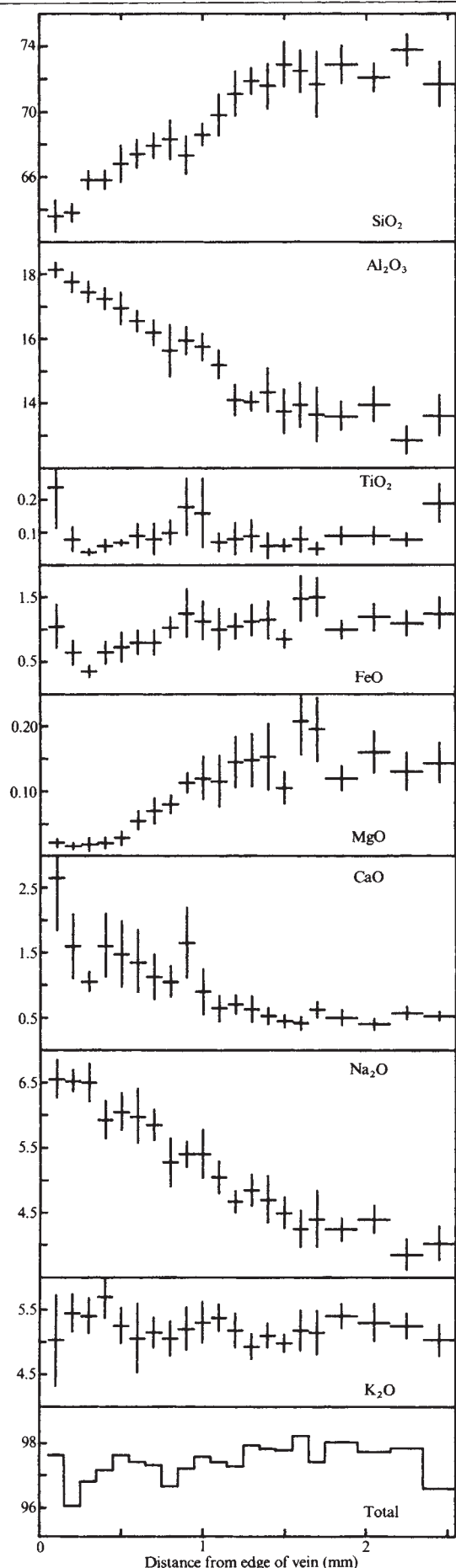


Fig. 3 Profiles of wt% oxides for whole-rock analyses of 100 μm wide strips, against distance from vein. Vertical bars, 2 standard errors on mean.

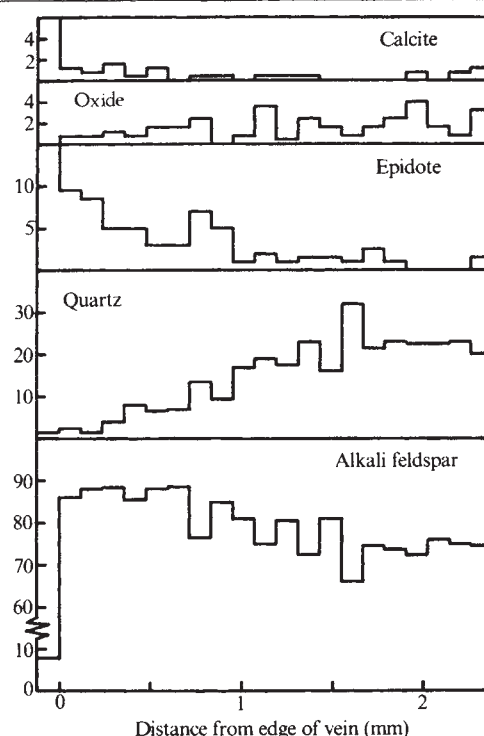


Fig. 4 Modal mineral percentages as a function of distance from vein.

slight enrichment of SrO as the vein is approached here (Table 1) shows that feldspar successfully incorporated Sr released by glass devitrification. Growth of potash feldspar enabled similar incorporation of the heavy fission product barium (Table 1). In the most severely altered material near the vein, epidote is a major hydrothermal mineral phase. Studies of hydrothermal borehole mineralogy³⁰ indicate that conditions for epidote growth have been achieved in the natural environment at depths of only 450 m. Analyses of the epidote group mineral allanite, which has grown in hydrothermal conditions in Skye³¹, reveal massive concentrations of actinides, rare-earth elements and strontium. The large A sites in this mineral can readily accommodate and trap LIL radioactive elements released during glass devitrification. Thus a new stable hydrothermal mineralogy can develop which will incorporate radioactive elements. However, glass must retain its structural integrity long enough for this hydrothermal mineralogy to develop.

This reveals the deficiency of borosilicate glass; the solubility of the borate fraction is such that it is readily removed from the system, causing the glass to disintegrate, and the radioactive elements to be released uncontrollably. Hence a lower borate fraction in radioactive waste glass will clearly result in greater resistance to leaching. Because borates form a flux for the melting process, their removal will necessitate higher furnace temperatures. However, if a colloidal solution of nitrates, with components present in similar proportions to the rhyolite eutectic, is evaporated, then dry melting should be possible at little over 1,000 °C (ref. 32).

That rhyolite glass retains its structural integrity and displays low solubility rates even when subject to devitrification and prolonged high temperature hydrothermal leaching indicates that high-silica glass is more suitable for long term disposal of radioactive wastes than borosilicate glass. Although high-silica glass requires a greater furnace temperature for melting, the extra cost involved may be outweighed by its stability advantages over borosilicate glass.

I thank R. A. Exley for advice in the field, S. Moorbath for helpful comments, the late R. Holland for preparation of the polished section, and P. Jackson and C. Fagg for advice on SEM and microprobe techniques. Microprobe research at Oxford is

supported by the NERC and I thank University College Oxford for a Weir Junior Research Fellowship, and the Department of Geology and Mineralogy for financial support.

Received 8 June; accepted 14 October 1981.

1. International Nuclear Information System (INIS) Atomindex Vol. 11, Part 5, 305 (Subject index of abstracts on vitrification) (1980).
2. *Symp. on Management of Radioactive Wastes from the Nuclear Fuel Cycle* (IAEA and NEA (OECD), Vienna, 1976).
3. Culler, F. Jr (ed.) *Proc. Int. Symp. Management of Waste from the LWR Fuel Cycle* (U.S. Energy Research and Development Administration, 1976).
4. Johnson, K. D. B. & Marples, J. A. C. *UKAEA, Harwell, AERE-R-9417* (HMSO, London, 1979).
5. Roberts, L. *Atom* **267**, 8–20 (1979).
6. Feats, F. & Keen, N. *New Scientist* **77**, 426–428 (1978).
7. Grover, J. R., Hardwick, W. H., Gayler, R. & Delve, M. H. *The FINGAL Process* (Conf-660208 (USAE), Richland, 1966).
8. Daglish, J. *Atom* **269**, 58–62 (1979); **290**, 306–311 (1980).
9. Chotin, M. M., Bonniaud, R. A., Jouan, A. F. & Rabot, G. E. in *Int. Symp. on Ceramics in Nuclear Waste Management* (eds Chikalla, T. D. & Mendel, J. E.) 73–81 (American Ceramic Society, Cincinnati, 1979).
10. Stone, J. A., Goforth, S. T. & Smith, P. K. A. *Meet. Am. ceram. Soc.* (Chicago, 1980).
11. Morris, J. B. *et al. Nature* **273**, 215–216 (1978).
12. *Kärn-Bränsle-Säkerhet Report* (AB Teleplan, SOLNA, Sweden 1978).
13. Merritt, W. F. *Proc. Symp. Management of Radioactive Wastes from Nuclear Fuel Cycle 2*, 27–35 (IAEA, Vienna 1976).
14. Strathdee, G. G., McIntyre, N. S. & Taylor, P. *Proc. int. Symp. Ceramics in Nuclear Waste Management* (Cincinnati, CONF-790420 243–247, U.S. Department of Energy, Oakridge, 1979).
15. Karkhanis, S. N., Bancroft, G. M., Fyfe, W. S. & Brown, J. D. *Nature* **284**, 435–437 (1980).
16. Gass, I. G. & Thorpe, R. S. 'Igneous Case Study' in *Science: A Third Level Course. Earth Science Topics and Methods* (ed. Aprehman, F.) (Open University Press, 1976).
17. Bott, M. H. & Tuson, J. *Nature* **242**, 114–116 (1973).
18. Forester, R. W. & Taylor, H. P. *Am. J. Sci.* **277**, 136–177 (1977).
19. Taylor, H. P. & Forester, R. W. *J. Petrol.* **20**, 355–419 (1979).
20. Norton, D. & Taylor, H. P. *J. Petrol.* **20**, 421–486 (1979).
21. Exley, R. A. thesis, Univ. Oxford (1980).
22. Brown, G. M. *Min. Mag.* **33**, 533–562 (1963).
23. Le Maitre, R. W. *J. Petrol.* **17**, 589–637 (1976).
24. Puchelt, H. *Handbook of Geochemistry* (ed. Wedepohl, K. H.) 56D (Springer Berlin, 1978).
25. McCarthy, G. J. *et al. Nature* **273**, 216–217 (1978).
26. Jenks, G. H. in *NWTS Conf. on Waste-Rock Interactions* (U.S. Energy Research and Development Administration, Report Y/OW1/SUB-77/14268, 5016, 1977).
27. Mendel, J. E. *Nucl. Technol.* **32**, 72–87 (1977).
28. Dickinson, A. P. & Exley, R. A. *Contr. Miner. Petrol.* **76**, 98–108 (1981).
29. Dickinson, A. P., Exley, R. A. & Smith, B. M. *Earth planet. Sci. Lett.* **51**, 58–70 (1980).
30. Tómasson, J. & Kristmannsdóttir, H. *Contr. Miner. Petrol.* **36**, 123–134 (1972).
31. Exley, R. A. *Earth planet. Sci. Lett.* **48**, 97–110 (1980).
32. Schairer, J. F. *J. Geol.* **58**, 512–517 (1950).

Fission track dating of zeolites

Sohan L. Koul & Lewis T. Chadderton

Division of Chemical Physics, CSIRO PO Box 160, Clayton, Victoria, Australia 3168

C. Kent Brooks

Institute of Petrology, University of Copenhagen, Øster Voldgade 5, 1350 Copenhagen K, Denmark

The fission track method has become acceptable in geochronology, not only in straightforward dating of igneous and metamorphic events but also in stratigraphic studies by the dating of ash layers^{1,2}, the elucidation of crustal uplift history^{3–7}, and in archaeology⁸. Such studies have mostly used accessory minerals such as apatite, sphene and zircon, whose track-retentive properties are well documented. However, fission tracks can be developed and observed under the optical microscope in most minerals^{9–11}. We report here on reconnaissance studies of fission track dating of zeolites, a complex group of tectosilicates, which is, for example, widely developed in basaltic rocks after their emplacement by later stage processes. We describe the etching conditions for some zeolites, in particular the minerals chabazite, stilbite and heulandite¹² and detail the track recording and retention characteristics for chabazite, with particular reference to fossil hydrothermal systems in the Faeroe Islands.

Fleischer *et al.*¹³ have previously reported the general etching conditions for zeolites. We have particularly investigated the effect of NaOH and KOH on chabazites. Homogeneous, inclusion-free chabazite crystals were annealed at 540 °C for 2 h

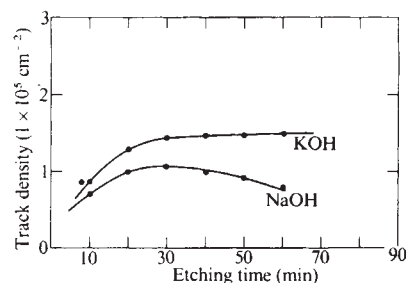


Fig. 1 Etching rates of chabazite for NaOH and KOH.

to erase all spontaneous tracks, mounted in epoxy resin and mechanically polished. The chabazite crystal structure only breaks down at 840 °C (ref. 14). Artificial tracks were then created using a ²⁵²Cf source in a 2π geometry and the chemical etching was carried out in two groups for various times using boiling NaOH and KOH. The results (Fig. 1) indicate that with NaOH a maximum is reached, followed by a decrease in track density due to the progressive removal of surface layers of changed crystal. In the case of KOH (some typical etch pits are shown in Fig. 2) a plateau, reached after a relatively short time, remains constant. These relationships are due to differences in the velocity of etching along the tracks (V_t) relative to that for the bulk solid (V_g). The ratio V_t/V_g is high for KOH but is much lower for NaOH. Note that uranium in chabazite is generally not considered to be an exchangeable ion.

Because thermal events may affect the final fission track 'age' of a sample we have carried out experiments both to evaluate this effect and to provide a calibration curve: chabazite is used as an example. If samples containing fission damage trails are subject to some thermal event and thereby lose a certain fraction of the tracks, they will appear to have decreased ages. Strozer and Wagner¹⁵ proposed a method of correction for thermally lowered track densities, using calibration curves of residual track density as a function of average length of etch pit, both quantities being expressed as fractions of their unannealed values^{16–18}. After annealing a chabazite at 520 °C for 6 h induced tracks were created at a thermal neutron dose of 10¹⁵ (nvt). The fragment track density was then determined by etching aliquots which had undergone heat treatments varying from 100 to 500 °C (±5 °C) for 1 h. Etching through cleavage [101̄] and [1̄101] planes was used to reveal the full track lengths. The length was evaluated by measuring both track dip and projected length, and phase contrast methods were used to enhance precision. An average of 'major' and 'minor' axes for each etch pit was recorded as the corresponding 'etch pit length'. For almost 11,000 etch pits scanned, between 800 and 900 track lengths were measured. Figure 3a shows the relationship between reduction in track density and length as a function of

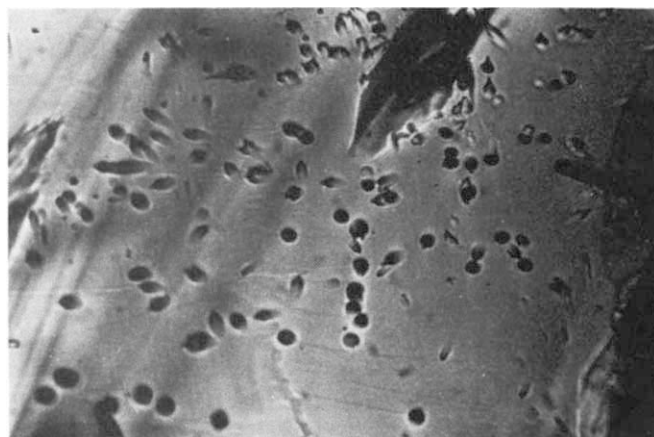


Fig. 2 Spontaneous fission tracks in chabazite etched in KOH (3 g in 4 g H₂O).

Table 1 Fission track results of minerals from the Faeroe Islands

No.	Sample location*	ρ_s (cm ⁻²)	ρ_i (cm ⁻²)	Dose (n)	<i>N</i>	<i>r</i>	Age (Myr ± 1σ)
Chabazite							
1	Runavik (US)	1.58 × 10 ⁴	1.72 × 10 ⁵	7.52 × 10 ¹⁵	4	0.995	42.5 ± 1.0
	Eysturoy	(464)	(625)				
1‡	Runavik	1.58 × 10 ⁴	1.35 × 10 ⁵	5.84 × 10 ¹⁵	6	0.989	42.1 ± 1.7
	Eysturoy (US)	(810)	(2,071)				
2	Slaettafjall	1.70 × 10 ⁴	3.44 × 10 ⁴	1.51 × 10 ¹⁵	6	0.959	45.9 ± 1.0
	Eysturoy (US)	1,436	2,309				
3	Malinsfjall	1.52 × 10 ³	1.34 × 10 ⁴	5.84 × 10 ¹⁵	5	0.954	40.8 ± 1.3
	Vidoy (US)	(430)	(1,295)				
3‡	Malinsfjall	1.51 × 10 ³	2.19 × 10 ⁴	9.80 × 10 ¹⁵	5	0.938	41.6 ± 1.1
	Vidoy (US)	550	1,667				
4	Nordara	1.75 × 10 ⁴	3.79 × 10 ⁴	1.51 × 10 ¹⁵	6	0.965	42.9 ± 1.0
	Sandoy (US)	1,102	1,696				
5	Eidi	7.14 × 10 ⁴	1.30 × 10 ⁵	1.42 × 10 ¹⁵	6	0.958	48.0 ± 1.0
	Eysturoy (MS)	1,421	2,286				
6	Tvöroyri	6.08 × 10 ⁴	3.05 × 10 ⁵	4.23 × 10 ¹⁵	5	0.679	51.9 ± 2.0
	Suduroy (LS)	179	1,565				
Stilbite							
7	Runavik	6.30 × 10 ⁴	1.28 × 10 ⁵	1.51 × 10 ¹⁵	7	0.989	45.7 ± 1.2
	Eysturoy (US)	350	2,488				
8	Gassá-Breida	1.63 × 10 ⁴	5.43 × 10 ⁴	2.48 × 10 ¹⁵	8	0.934	45.8 ± 0.9
	Streymoy (US)	883	1,565				
9	Sörvagur	2.87 × 10 ⁴	2.44 × 10 ⁵	6.24 × 10 ¹⁵	6	0.754	45.2 ± 1.7
	Vagar† (MS)	2,150	2,200				
9‡	Sörvagur	2.87 × 10 ⁴	2.77 × 10 ⁵	7.12 × 10 ¹⁵	8	0.943	45.4 ± 1.9
	Vagar (MS)	(234)	(1,112)				
10	Harnar	4.35 × 10 ⁴	1.26 × 10 ⁵	2.49 × 10 ¹⁵	6	0.931	52.9 ± 2.1
	Suduroy (LS)	367	1,384				
10‡	Harnar	4.36 × 10 ⁴	3.63 × 10 ⁵	7.12 × 10 ¹⁵	6	0.892	52.6 ± 2.2
	Suduroy (LS)	(201)	(519)				
11	Frodböur	4.56 × 10 ⁵	7.65 × 10 ⁵	1.51 × 10 ¹⁵	8	0.768	55.4 ± 2.5
	Suduroy (LS)	240	533				
Heulandite							
12	Smorbushellisgjogv	4.96 × 10 ⁴	4.52 × 10 ⁵	8.04 × 10 ¹⁵	8	0.973	54.3 ± 2.0
	Mykines (LS)	189	1113				

ρ_s is the spontaneous track density; ρ_i the induced track density; *n*, the thermal neutron dose (nvt); *N* the number of grains counted; *R* the correlation coefficient.

* US, Upper series; MS, middle series; LS, lower series.

† Fissure-filling, all others are from amygdalites.

‡ Muscovite detector used for neutron dose determination. (For further details see text.)

temperature. This information gives an immediate estimate of the degree of fading of the spontaneous tracks and hence the age reduction provided that the track lengths are measured. Chabazite has a measured average track length of $11.8 \pm 2.1 \mu\text{m}$. At 100°C , a slight decrease of $\sim 2.8\%$ in the track density and $\sim 3.5\%$ in the mean track length of etch pits was observed. At higher temperatures, the track length decreased more rapidly than the track density. After 1 h at 500°C constant temperature the track density was reduced by 57.3% and the track length by 78.3%. After 1 h at 600°C all the tracks were annealed out. The reduction in track density was observed to lag behind that of track length (Fig. 3a). U^{235} -induced fission tracks in both unannealed and annealed samples were measured. The shrinkage in track length, plotted as a function of track density reduction, yields the appropriate calibration relationship. The ratio of mean etch pit length of annealed and unannealed fission tracks is shown as a function of the percentage of fission track density reduction in Fig. 3b. The resulting calibration curve has been used to correct thermally-lowered fission track ages of chabazite (estimated correction is $\sim 3\%$). Studies of various etchants for stilbite and heulandite resulted in the use of 2% HF at 23°C for 20–30 s and 10 ml aqua regia: 1 ml 2% HF at 23°C for 30–50 s respectively for these two minerals.

The zeolite specimens were dated by the (EDM) detector method¹⁹, samples of chabazite, stilbite and heulandite being sandwiched between muscovite mica detectors for measuring the induced tracks. The integrated thermal neutron dose to which the samples were exposed was itself determined by irradiating calibrated glass dosimeters with each consignment of

zeolites. Each glass dosimeter (supplied by R. Fleischer) was fractured and the fresh surfaces were etched with newly prepared 20% HF for ~ 30 – 35 s at 23°C to obtain well defined, easily distinguished etch pits, ~ 800 – $1,000$ pits being scanned for each consignment. The surface density of etch pits was measured and the integrated neutron flux calculated using the calibration relation^{20,21}, and a factor of 2.26×10^{11} neutrons per track. Sample numbers 1, 3, 9 and 10 were again dated where muscovite detector was used for neutron dose measurement. Thermal neutron dose was determined by counting tracks in muscovite detector irradiated in contact with standard NBS glasses SRM-962 and SRM-963 (Cu values)²²; the values obtained agreed with the independent measurements made on Au foil. The analytical results are in accord with the age results given in Table 1.

An investigation of the general relationship between track counting on internal mineral surfaces and that on external detectors was also made to determine the geometric factor²³ for the use of external detectors with zeolites. For the zeolites studied the geometric factor varied between 0.51 ± 0.02 and 0.57 ± 0.03 , assuming the etching rate and the counting efficiency for external detectors to be the same for each zeolite, and the track shape to be virtually the same. However, etching of tracks lying in different orientations was highly anisotropic, giving rise to etch pits of varying topographic appearance. The constants used in the age calculations were $\lambda_t = 6.8 \times 10^{-17} \text{ yr}^{-1}$, $I = 7.223 \times 10^{-3}$, $\sigma = 5.802 \times 10^{-22} \text{ cm}^2$ and $\lambda_d = 1.551 \times 10^{-10} \text{ yr}^{-1}$. Errors were calculated according to the method of McGee and Johnson²⁴ and Naeser *et al.*²⁵, expressed in 1 s.d.

The exposed part of the Faeroe Islands consists of up to 3 km of largely basaltic lavas divided into a lower, middle and upper series by stratigraphic marker horizons. There are distinct mineralogical and geochemical differences throughout the pile (see refs 26–29). The lithologies show that significant time gaps separated the three series, but this has not shown up in radiometric dating, where variable ages between 49.2 ± 1.5 – 60.4 ± 1.4 Myr were obtained³⁰. Subsequent refinement of these data³¹ suggests that the Faeroese lava pile from the bottom of the lower to the top of the middle series was erupted in the interval 55.2 ± 1.0 – 54.6 ± 1.2 Myr ago.

Fission track results for chabazite, stilbite and heulandite are shown in Table 1 for samples shown in Fig. 4 where a spread of ages from 40.8 ± 1.3 to 55.4 ± 2.5 Myr have been found. The oldest values occur in the lower series and the youngest in the upper series. However, we have no other evidence to contradict the K–Ar results which indicate a relatively short time span for the volcanic activity, consistent with results from the contemporaneous East Greenland and lavas³². The observed distribution is probably regional, with the youngest ages in the northeastern part of the islands, reflecting a more prolonged cooling in that region. Ages from the lower series on Suduroy and Mykines, which are close to the accepted age of volcanism, indicate that in the south and west of the islands the lavas underwent rapid cooling either as a result of lack of burial by significant amounts of overlying material, or as a result of a very rapid uplift. Such rapid uplift is probably due to the dome-shaped structures believed to be present offshore to the north-west of Mykines and under Suduroy³³. These domes, although occurring on a much smaller scale, invite comparison with East

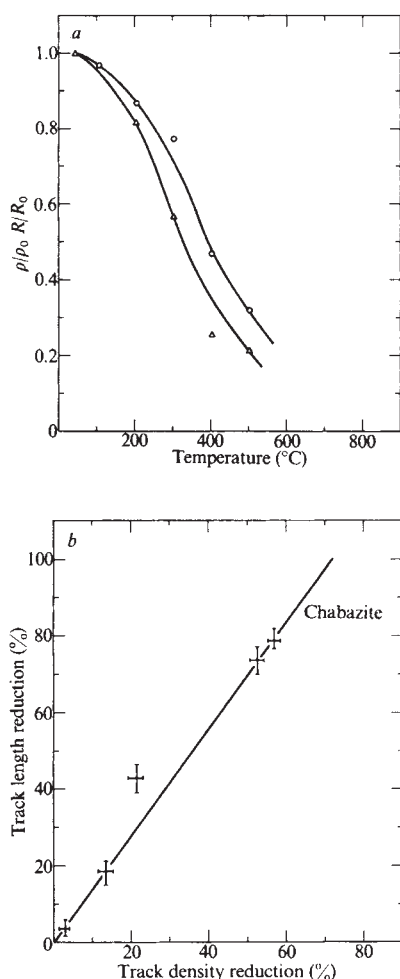


Fig. 3 a, R/R_0 (Δ) and ρ/ρ_0 ($\circ$) as a function of temperature T for chabazite. b, Percentage reduction of fission track density versus track length in chabazite.

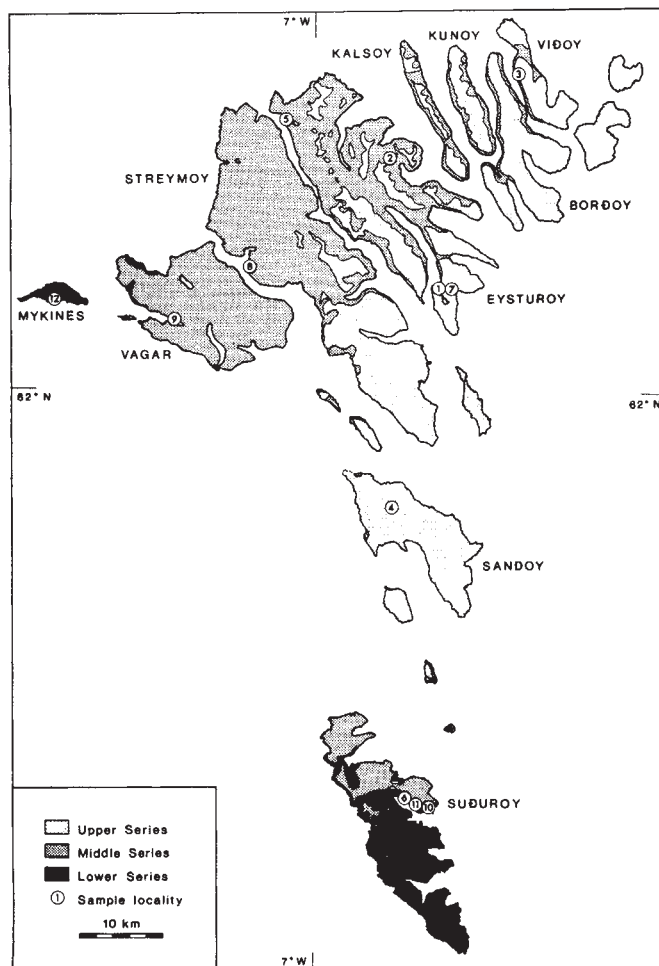


Fig. 4 Geological map of the Faeroe Islands after ref. 26.

Greenland, where a large domal uplift³⁴ was raised shortly after the basaltic volcanism and rapidly dissected by erosion⁴.

The annealing experiments reveal that the thermal stability of tracks in chabazite is lower than that in sphenes, garnet, epidote, allanite and hornblende³⁵. Data from both annealing experiments and track shrinkage suggest that ages determined by applying the fission track method to zeolites will be slightly affected by thermal annealing. We assume in fact that the tracks became stable rather suddenly. Nevertheless, our study certainly indicates that the fission track method is suitable for zeolite dating. Work on the Faeroese zeolites will continue to try to clarify the detailed history of the lava pile.

We thank Ole Jørgensen for samples, and Dr R. L. Fleischer, General Electric Research Laboratory, Schenectady, New York, for supplying the glass dosimeter.

Received 19 June; accepted 28 September 1981.

1. Naeser, C. W., Izett, G. A. & Wilcon, R. E. *Geology* **1**, 93 (1973).
2. Gleadow, A. J. W. *Nature* **284**, 225 (1980).
3. Harrison, T. M., Armstrong, R. L., Naeser, G. W. & Harakal, J. E. *Can. J. Earth Sci.* **16**, 400 (1979).
4. Gleadow, A. J. W. & Brooks, C. K. *Contr. Miner. Petrol.* **71**, 45 (1979).
5. Doherty, J. T. & Lyons, J. B. *Bull. geol. Soc. Am.* **91**, 16 (1980).
6. Wagner, G. A. *Earth planet. Sci. Lett.* **4**, 411 (1968).
7. Wagner, G. A. & Reimer, G. M. *Earth planet. Sci. Lett.* **14**, 263 (1972).
8. Fleischer, R. L., Price, P. B. & Walker, R. M. *Science* **148**, 72 (1965).
9. Price, P. B. & Walker, R. M. *J. geophys. Res.* **68**, 4847 (1963).
10. Koul, S. L. & Virk, H. S. *Geophys. Res. Bull.* **14**, 139 (1976).
11. Koul, S. L. *Rad. Effects Lett.* **43**, 7 (1979).
12. Deer, W. A., Howie, R. H. & Zussman, J. *Rock Forming Minerals* Vol. 4, 351 (Longmans, London, 1963).
13. Fleischer, R. L., Price, P. B. & Walker, R. M. *Nuclear Tracks in Solids* (University of California Press, London, 1975).
14. Sutherland, F. L. J. *J. geol. Soc. Austr.* **24**, 171 (1977).
15. Strozer, D. & Wagner, G. A. *Earth planet. Sci. Lett.* **5**, 463 (1960).
16. Wagner, G. A. & Strozer, D. *Eclogae. geol. Helv.* **63**, 335 (1970).
17. Strozer, D. *Earth planet. Sci. Lett.* **8**, 55 (1970).
18. Khan, H. A. & Durrani, S. A. *Radiat. Eff.* **13**, 257 (1973).
19. Gleadow, A. J. W. & Lovering, J. F. J. *J. geol. Soc. Austr.* **25**, 323 (1978).

20. Fleischer, R. L., Price, P. B. & Wilker, R. M. *Nucl. Sci. Engng* **22**, 153 (1965).
21. Fleischer, R. L. & Hart, H. R. Jr *Fission Track Dating: Techniques and Problems*, 139 (Scottish Academic, New York, 1972).
22. Carpenter, B. S. & Reimer, G. M. *Natn. Bur. Standards Spec. Publ.* 260-49, 16 (1974).
23. Gleadow, A. J. W. & Lovering, J. F. *Nucl. Track. Detection* **1** (1977).
24. McGee, V. E. & Johnson, N. M. *Math. Geol.* **11**, 255 (1979).
25. Naeser, C. W., Johnson, N. M. & McGee, V. E. *U.S. geol. Surv. Open-File Rep.* **78**, 303 (1978).
26. Rasmussen, J. & Noe-Nygaard, A. *Geol. Surv. Denmark* **1**, 25 (Copenhagen).
27. Waagstein, R. thesis, Univ. Copenhagen (1977).
28. Bollingberg, H., Brooks, C. K. & Noe-Nygaard, A. *Bull. geol. Soc., Denmark* **24**, 55 (1975).
29. Schilling, J. G. & Noe-Nygaard, A. *Earth planet. Sci. Lett.* **24**, 1 (1974).
30. Tarling, D. H. & Gale, N. H. *Nature* **218**, 1043 (1968).
31. Fitch, F. J., Hooker, P. J., Miller, J. A. & Brereton, N. R. *J. geol. Soc. Lond.* **135**, 499 (1978).
32. Soper, N. J., Downie, C., Higgins, A. C. & Costa, L. I. *Earth planet. Sci. Lett.* **32**, 149 (1976).
33. Schröder, N. F. *Ann. Soc. Scient. Faeroensis* **19**, 20 (1971).
34. Brooks, C. K. *Meddr. Grønland, Geosci.* **1**, 1 (1979).
35. Saini, H. S., Sharma, O. P. & Nagpaul, K. K. *Nucl. Tracks* **3**, 139 (1979).

Comb-layering in carbonatite dykes

Kerstin Katz & Jörg Keller

Mineralogisch-Petrographisches Institut, Universität Freiburg, FRG

Carbonatitic dykes from the Tertiary Kaiserstuhl volcano in the Upper Rhinegraben, FRG, commonly show fine-banded layering with skeletal branching calcite crystals oriented perpendicular to the layering. These calcite comb-layers are produced from magmatic *in-situ* crystallization in specific conditions of supercooling and rapid cooling rates^{1,2}. We show here that the crystallization features underline the magmatic origin of the observed textures, mineral phases and melt compositions and are a key to the physicochemical nature of the carbonatitic magma.

The term 'comb-layering' has been applied to textures in magmatic rocks characterized by layers of elongate, parallel and branching crystals with preferential orientation perpendicular to the plane of layering³. It is mostly feldspars, pyroxenes, olivine and amphibole that form the elongate crystals of comb-layers in magmatic intrusions. Moore⁴ found comb-layers of perovskite in olivine-melilitite and Dawson (unpublished data) observed sphene comb-layers in an ijolithe block from Oldoinyo Lengai. One example of elongate and branching calcite has been reported from a kimberlite sill at Benfontein, South Africa⁵. Calcite dendrites have been found in some natural carbonatites^{6,7} and in calcite melting experiments^{8,9}.

In general, natural examples of comb-layering have been explained as the result of crystallization from a supercooled magmatic melt. Crystallization experiments have reproduced these comb-layer textures² and confirmed that specific conditions of supercooling, supersaturation and cooling rates are responsible for their formation.

Carbonatites in the Tertiary Kaiserstuhl alkaline complex occur as three different rock types¹⁰⁻¹²: (1) a central intrusion of coarse-grained sövite; (2) several hundred dykes cutting the subvolcanic and volcanic silicate rocks; and (3) carbonatitic lapillistones and ashes formed by the surface eruption of the carbonatite magma. The carbonatitic dykes (<1 cm to 1 m wide) exhibit a wide variety of textures from trachytoid and skeletal to porphyritic and granular. Most of the dykes consist of >90 vol.% of calcite; only a few contain larger amounts of silicate phases. Magnetite, apatite and phlogopitic mica are the main non-carbonate minerals. All dykes are alvikitic according to the nomenclature of Le Bas¹³.

Comb-layering parallel to the dyke margins has been observed in many of these dykes. The arrangement of layers is very regular, with a well-defined median symmetry plane and pairs of corresponding layers on either side of the centre (Figs 1, 2). The layers consist of elongate calcite crystals with a strictly parallel orientation perpendicular to the layering. Narrow dykes (maximum width 10 cm) may consist of one pair of symmetrical comb-layers only ('simply-layered dykes') (Fig. 1). Wider dykes

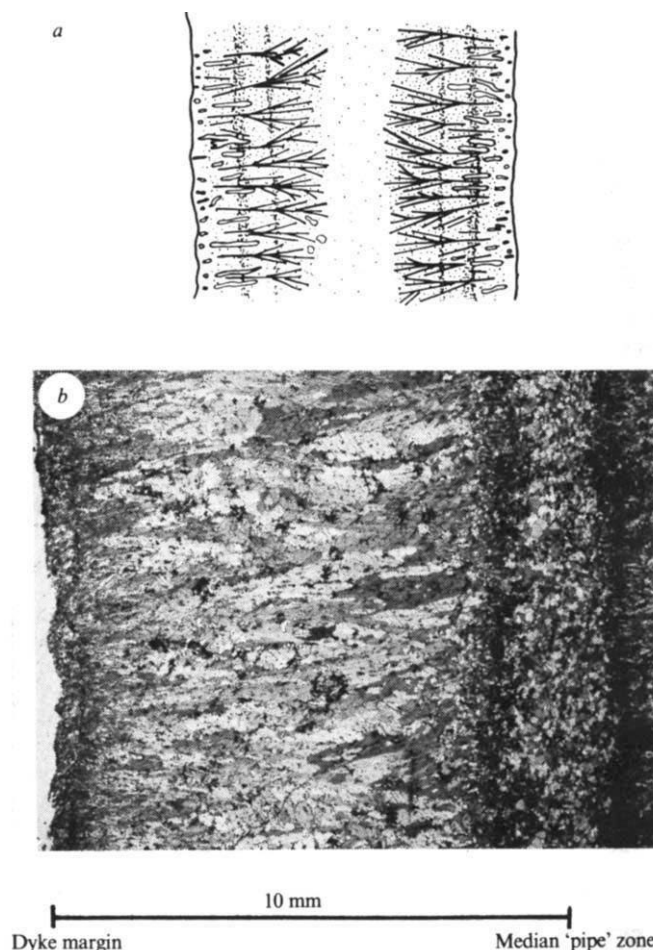


Fig. 1 *a*, Cross-section through simply-layered carbonatite dyke (width 5 cm) with one pair of symmetrical comb-layers formed by skeletal calcite dendrites. Border vesicle zones and a fine-grained central zone are also developed. *b*, Micrograph of cross-section of a simply-layered carbonatite dyke with a comb-layer of columnar calcite and a granular central 'pipe' zone.

sometimes show a rhythmic banding with a succession of several comb-layers, for example Fig. 2. The layers are separated by thin bands of fine-grained carbonatite. A zone of aphyric, very fine-grained carbonatite which separates comb-layers from the actual wall-rock contact is interpreted as the chilled margin.

Single comb-layers can be up to 3 cm wide or considerably smaller (<1 mm). The elongate calcite crystals reach a length of ~2 cm with a maximum length-to-width ratio of 150:1. The elongation is perpendicular to the *c*-axis. Mostly the crystals are skeletal and show dendritic bifurcations (Fig. 3). The dendrites always branch towards the centre of the dykes where the elongate comb-layer calcites become closely spaced forming fan-shaped bundles. A curvature of single dendritic and skeletal calcite crystals is common (Fig. 3*a*). The interstitial matrix between the skeletal crystals consists of granular or trachytoid calcite with extremely fine-grained (<30 μ m) acicular apatite

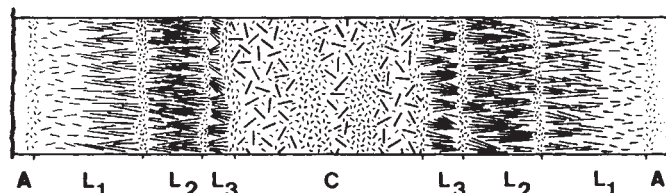


Fig. 2 Schematic cross-section through rhythmically layered comb-layer carbonatite dyke (width 25 cm) A, aphyric border; L₁₋₃, symmetrical pairs of comb-layers; C, microporphyritic to equigranular central zone.

and euhedral magnetite in varying amounts. Sometimes tiny pyrochlore octahedra may be present.

The mineral chemistry¹² of calcite is rather uniform with very little variation between different dykes or single layers in one dyke. It is pure calcite with $\sum \text{MnCO}_3 + \text{FeCO}_3 + \text{MgCO}_3 < 0.5 \text{ wt\%}$. The SrCO_3 content is high and averages to 2 wt%. Apatite chemistry is characterized by high rare earth element contents (average La_2O_3 0.5, Ce_2O_3 1.0 wt%) and significant SiO_2 contents (up to 3.0 wt%). Low fluorine concentrations (average 0.75 wt%) are in marked contrast to higher F values in apatites of the Kaiserstuhl sövites (average 2 wt%).

The granular central part between the comb-layers varies in thickness from a few millimetres in simply-layered dykes (Fig. 1) to several centimetres (Fig. 2). The contact between the centre and the comb-layer is often rather sharp. Some of the larger central zones show flow structures and in a few cases contain detached pieces of adjacent comb-layers. Vesicle zones parallel to the border occur in a few dykes. The vesicles may be void but are often filled with pure calcitic material. In the comb-layers these vesicles are elongate parallel to the dendrites. In some dykes they show a striking similarity to the carbonatitic 'migra-

tion vesicles' described by Dawson and Hawthorne⁵ from the kimberlite sill at Benfontein. In two dykes apatite comb-layers occur as well as the calcite ones. The apatite crystals are elongate and skeletal and branch towards the dyke centre.

Simply and rhythmically layered dykes have been chemically analysed (Table 1). Their composition shows the characteristic trace-element enrichment (for example, Sr, Ba, Nb, light rare-earth elements), which is typically carbonatitic and compares well with other carbonatitic rocks of the Kaiserstuhl. There is a close chemical similarity with rapidly quenched volcanic lapilli which occur in the form of carbonatite lava droplets in the Kaiserstuhl (Table 1, no. 5). With H. W. Hubberten (unpublished data) we have studied carbon- and oxygen- isotope ratios of carbonatites from the Kaiserstuhl and analysed several comb-layer zones. Their isotope ratios are within the narrow limits of primary carbonatites^{14,15}. Values as low as -5 to -7% for $\delta^{13}\text{C}$ (PDB) and $+5$ to $+8\%$ for $\delta^{18}\text{O}$ (SMOW) have been determined.

The conditions necessary for comb-layer formation have been discussed by Donaldson², briefly they are: a compositional or thermal gradient causing unidirectional solidification; a degree of supercooling at which homogeneous nucleation in the melt is low and crystal growth is fast; a cooling rate that allows isotherms to advance at the same rate as the elongate crystal growth rate.

The almost monomineralic nature of the carbonatite melt reduces the importance of compositional factors on the constrained crystal growth of calcite. In this respect carbonatite magmas more closely resemble metal melts than multi-component silicate magmas. The small, simply-layered carbonatite dykes therefore may be regarded as natural analogues of metal casts with 'columnar' wall-zones and a central 'pipe' zone (Fig. 1b). The main reason, therefore, for directed crystal growth in the narrow carbonatite dykes is supercooling in a high thermal gradient perpendicular to the dyke walls. The presence of comb-layering with elongate calcite in very narrow dykes ($<1 \text{ cm}$) shows a high growth rate of calcite at rapid cooling. If the cooling rate exceeds the growth rate, supercooling in the

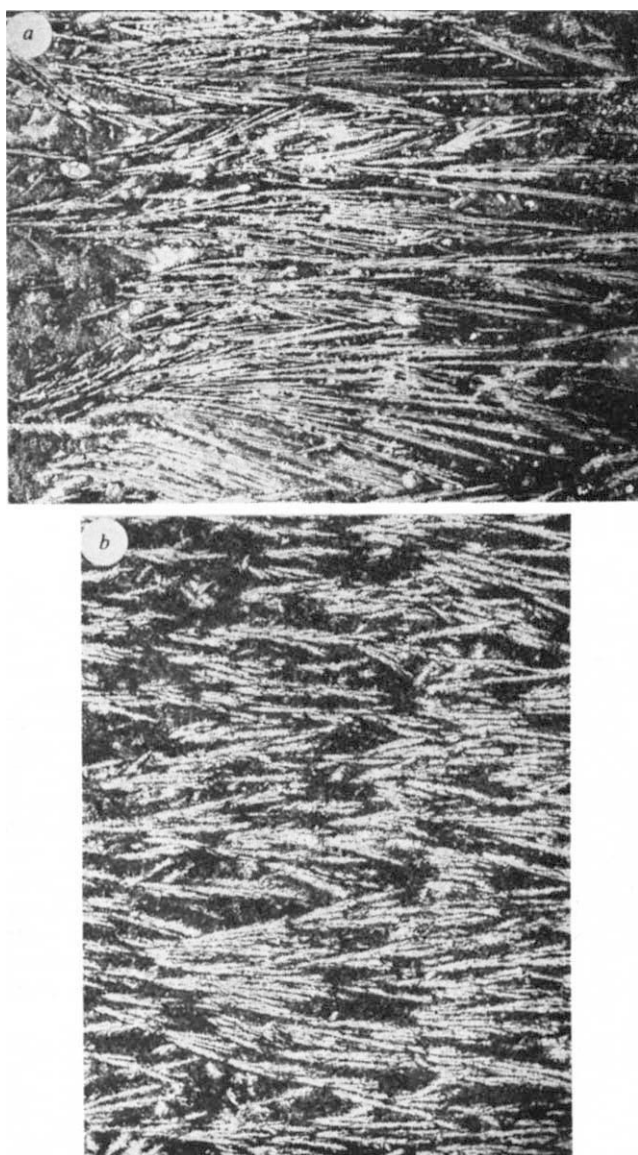


Fig. 3 *a*, Skeletal and dendritic calcite crystals with a slight curvature forming comb-layer of a carbonatite dyke (photographed area: $8.5 \times 6.5 \text{ mm}$). *b*, Small bundles of calcite dendrites branching towards the centre of the carbonatite dyke. The dark matrix between the dendrites is rich in extremely fine-grained magnetite and apatite (photographed area: $3.25 \times 2.5 \text{ mm}$).

Table 1 Chemical composition of comb-layered carbonatite dykes. Comparison with other carbonatite rocks from the Kaiserstuhl shows the typical chemical features of carbonatites

	1	2	3	4	5
SiO_2	2.22	1.74	1.69	1.37	1.02
TiO_2	0.10	0.08	0.09	0.05	0.05
Al_2O_3	0.58	0.66	0.48	0.25	0.36
$\text{Fe}_2\text{O}_{3\text{tot}}$	3.03	3.97	3.03	3.18	1.67
MnO	0.46	0.57	0.69	0.42	0.54
MgO	0.45	1.15	0.73	3.05	0.45
CaO	49.2	48.7	49.0	48.6	51.7
Na_2O	0.10	0.00	0.08	0.10	0.05
K_2O	0.05	0.00	0.09	0.03	0.02
P_2O_5	1.87	2.45	1.74	2.87	1.39
LOI	38.7	37.3	38.4	37.2	40.5
V	230	150	240	90	110
Sr	11,300	7,700	7,850	9,000	6,000
Y	100	100	60	75	25
Zr	45	140	60	180	15
Nb	1,030	1,020	810	2,400	670
Ba	1,220	740	1,800	650	1,200
La	1,000	910	640	450	350
Ce	1,850	1,770	1,100	1,100	560
Nd	400	440	230	260	110

Analyses were by X-ray fluorescence.

Column 1: Simply-layered carbonatite dyke, 3cm wide; Oberbergen (bulk sample).

Column 2: Rhythmically layered carbonatite dyke, 25cm wide; Schelingen. Bulk sample, single layers show no significant variation.

Column 3: Average of 51 carbonatite dykes from the Kaiserstuhl.

Column 4: Coarse-grained sövite; Schelingen.

Column 5: Carbonatite lapilli; Kirchberg, Niederrotteil.

magma ahead of the solidification front will soon reach a value at which homogeneous nucleation takes place and comb-layer formation stops². We conclude that comb-layered carbonatite dykes originated by intrusion of an almost nuclei-free melt in which heterogeneous nucleation only occurred at the dyke walls. A specific degree of supercooling maintained the low rate of homogeneous nucleation and elongate calcite crystals could grow along a thermal gradient in a rapidly cooling dyke. Experimental results¹ suggest that higher degrees of supercooling are required for comb-layer formation in melts of lower viscosity. Intrusive features of dykes and the nature of extrusive carbonatites of identical composition¹¹ suggest an extremely low viscosity of the carbonatite magma.

The origin of rhythmically layered carbonatite dykes is more complex. Compositional supersaturation can explain alternating comb-layers of differing mineralogy¹ and is suggested for the origin of the rare succession of calcite and apatite layers. Compositional supersaturation cannot be evoked for the sequence of comb-layers of very similar composition such as the calcite layers in carbonatite dykes. Periodic buildup and release of volatile pressure have been suggested as a cause for rhythmic banding^{1,2}. This seems plausible for the shallow subvolcanic to near-surface environment in which the dykes crystallized. As comb-layer crystallization proceeds the melt becomes enriched in volatiles. This causes depression of the liquidus and stops the growth of elongate and skeletal calcite. A boundary layer is formed along which a new comb-layer starts crystallizing when appropriate supercooling conditions are re-established.

The above interpretation of banded comb-layered carbonatite dykes gives important data on the nature of the carbonatite magma. The intruding melt was almost completely liquid and dendritic calcite crystals are evidence of primary textures. Calcite was the liquidus phase in the carbonatite magma. Secondary recrystallization is of very minor importance in the present examples. The rapid cooling in narrow dykes suggests that the composition of analysed carbonatites is a good approximation to the original composition of carbonatitic melt.

This study was supported by a grant from Deutsche Forschungsgemeinschaft. We thank J. B. Dawson and C. Donaldson for valuable comments.

Received 8 July; accepted 7 October 1981.

1. Lofgren, G. E. & Donaldson, C. H. *Contr. Miner. Petrol.* **49**, 309–319 (1975).
2. Donaldson, C. H. *Miner. Mag.* **41**, 323–355 (1977).
3. Moore, J. G. & Lockwood, J. P. *Bull. geol. Soc. Am.* **84**, 1–20 (1973).
4. Moore, A. A. *Miner. Mag.* **44**, 147–150 (1981).
5. Dawson, J. B. & Hawthorne, J. B. *J. geol. Soc. Lond.* **129**, 239–278 (1973).
6. Girault, G. *Bull. Soc. fr. Minér. Cristallogr.* **89**, 496–513 (1966).
7. Gittins, J., Hewins, R. H. & Laurin, A. F. *Phys. Chem. Earth* **9**, 137–148 (1975).
8. Paterson, M. S. *Am. Miner.* **43**, 603–606 (1958).
9. Wyllie, P. & Tuttle, O. F. *J. Petrol.* **1**, 1–46 (1960).
10. Wimmenauer, W. in *Carbonatites* (eds Tuttle, O. F. & Gittins, J.) 183–204 (Interscience, New York, 1966).
11. Keller, J. J. *Volcan. Geotherm. Res.* **9**, 423–431 (1981).
12. Katz, K. thesis Univ. Freiburg (in the press).
13. Le Bas, M. J. *Carbonatite-Nephelinite Volcanism*; London, (Wiley, London, 1977).
14. Hoefs, J. *Stable Isotope Geochemistry* (Springer, Berlin, 1973).
15. Sheppard, S. M. F. & Dawson, J. B. *Phys. Chem. Earth* **9**, 747–763 (1975).

Biologically damaging radiation amplified by ozone depletions

S. A. W. Gerstl & A. Zardecki

Theoretical Division, MS 210, Los Alamos National Laboratory,
Los Alamos, New Mexico 87545, USA

H. L. Wiser

Office of Research and Development, US Environmental Protection
Agency, Washington DC 20460, USA

Many recent studies^{1–3} indicate that releases of chlorofluorocarbons (CFCs)—mainly chlorofluoromethanes (CFMs)—into the atmosphere deplete the stratospheric ozone layer. Potentially dangerous consequences of these ozone depletions, such as increases in skin cancer, are expected due to a subsequent

increase in biologically damaging solar UV radiation reaching the ground. The biological effectiveness of this radiation amplification can be quantified by a radiation amplification factor (RAF) for which a value of 2 has been previously assumed: RAF=2 means, for example, that a 1% ozone depletion will result in a 2% increase in damaging UV dose at ground level. Using accurate radiative transfer calculations together with a detailed modern data base, we calculate here the RAFs for erythemally and DNA-weighted UV-B dose assuming ozone depletions as resulting from a two-dimensional model, as well as a global ozone depletion of 10%. The amplification factor as a function of latitude and season is found to be between 1.9 and 2.2 for erythema and between 2.5 and 2.8 for DNA. This is a smaller range variation than previously claimed.

In theoretical investigations of the problem using one-dimensional numerical models, the most probable value calculated for the eventual steady-state ozone depletion due to continued release of F-11 and F-12 at the 1977 level is 5–10%. New values of some of the chemical rate coefficients, such as the reaction rate of OH with HNO₃ (ref. 4), lead to this ozone depletion. Studies using two-dimensional models indicate that any change in ozone will not be uniform, but will show variations with latitude and season.

In estimating the resulting damaging UV dose rate (DUV) for erythema and DNA action spectra, we consider both a globally uniform ozone depletion consistent with the predictions of one-dimensional models, and latitude/season-dependent depletions as predicted by the two-dimensional model of Pyle⁵. Extensive calculations of this type have already been performed during the US Climatic Impact Assessment Program (CIAP) (see ref. 7). Our new calculations use a highly efficient and versatile radiative transfer code⁸ in which we incorporated a detailed data base for realistic atmospheres and which allows the computation of solar irradiance spectra to a predetermined degree of accuracy. Up to 70 km altitude we consider molecular and aerosol extinction and scattering profiles and include all

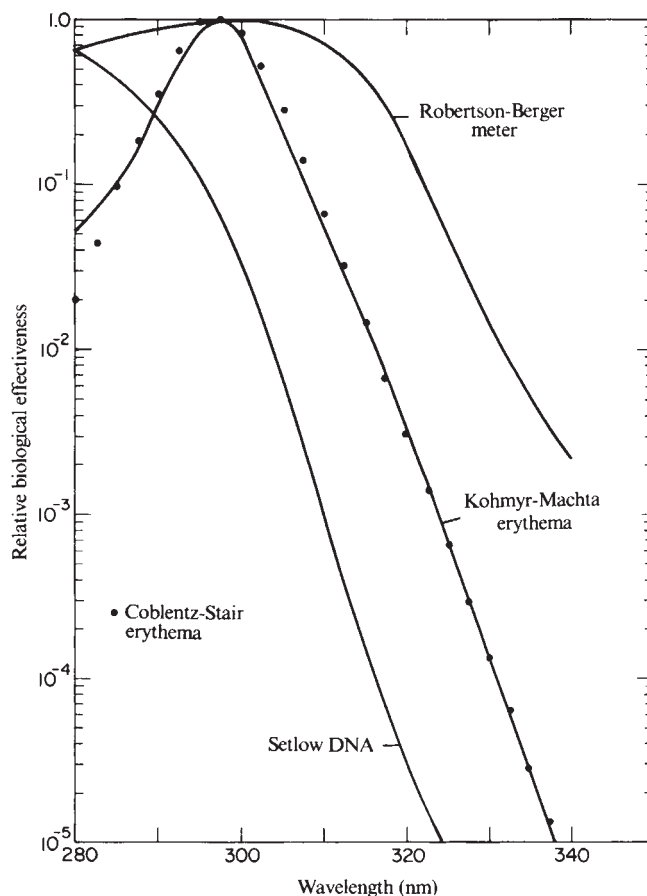


Fig. 1 Action spectra in use for biological effects of UV radiation.

orders of scattering in our computations. Our atmosphere model allows us to extend the results of Green and Mo⁷ to larger latitudes.

We calculate first the downward solar flux at ground level for the biologically effective solar UV spectral region from 0.28 to 0.34 μm (UV-B) for ozone depletions of 1–20%. Molecular absorption and scattering coefficients were prepared for five one-dimensional atmosphere models which correlate latitude and season⁹: (1) tropical; (2) midlatitude summer; (3) midlatitude winter; (4) subarctic summer; and (5) subarctic winter. Although the use of only five model atmospheres is obviously less detailed than a full two-dimensional model, this data base has proved to be adequate in extensive transmittance and radiance calculations as verified against laboratory and satellite measurements¹⁰. The wavelength dependency of the ozone absorption coefficient is described by an analytical representation also used by Shettle and Green¹¹. We include absorption and scattering by aerosols which are assumed to be distributed according to a rural model profile¹².

The biologically damaging effect due to UV-B radiation reaching the ground may be quantified by the DUV dose rate calculated from

$$\text{DUV} = \int_{0.28\mu\text{m}}^{0.34\mu\text{m}} E(\lambda) I(\lambda, Z_0) d\lambda$$

where $I(\lambda, Z_0)$ is the irradiance at ground level for the solar zenith angle Z_0 , as computed from a radiative transfer calculation. $E(\lambda)$ is the action spectrum characterizing a specific biological effect. Although changes in atmospheric ozone may produce rather small changes in the total solar flux reaching the ground, the changes in the UV-B part of the spectrum can be of considerable biological significance. The relative effectiveness of different wavelengths (the action spectrum $E(\lambda)$) for most biological UV effects increases with decreasing wavelength. This function $E(\lambda)$ must therefore be used as a weighting function for the solar irradiance when biological effects are to be estimated, as shown above for DUV. As can be seen from Fig. 1, the wavelength dependence of $E(\lambda)$ is significantly different for biological effects². Chemical compounds in biological systems that absorb UV-B radiation are mainly proteins and nucleic acids. For our calculations we chose as action spectrum for human sunburn the standard erythral action spectrum¹³ of Coblentz and Stair. The spectral sensitivity curve of the Kohmyr and Machta¹⁴ modified Dobson instrument, used by Pyle and Derwent⁶ as erythral weighting function, is very similar to the Coblentz–Stair action spectrum, as shown in Fig. 1. As the action spectrum for human skin cancer is not known, we use the generalized DNA action spectrum compiled by Setlow¹⁵ as a generic representation for this damaging UV-B effect. The stronger response to shorter wavelengths of Setlow's DNA action spectrum as compared with the erythral action spectrum leads to different results for the calculated DUV rates and RAFs. For completeness we also reproduce in Fig. 1 the spectral response curve of the Robertson–Berger meter which is used in the existing network of worldwide UV dose measurements. However, as seen from Fig. 1, and as emphasized in ref. 2, the response characteristics of this meter are quite different than the erythral and the DNA action spectrum.

As the solar irradiance at ground level depends on the solar elevation, an effective solar zenith angle $Z_0 = \cos^{-1} \langle \mu_0 \rangle$ is determined as a daily average over the continuously varying zenith angles at given date and latitude:

$$\langle \mu_0 \rangle = \frac{1}{T} \int_{t_{\text{sunrise}}}^{t_{\text{sunset}}} \mu_0(t) dt$$

where T is the duration of daylight. Figure 2 shows a plot of Z_0 against latitude and season. This allows us to correlate the calculated DUV dose for a given effective zenith angle Z_0 , and a given atmosphere model, with latitude and season. The consistency of using the effective solar zenith angle Z_0 has been checked against the procedure in which the DUV dose is taken at several zenith angles and then averaged over the day. The

daily DUV doses as obtained by these two different averaging methods differ by <5%, which we consider adequate for this investigation.

The RAF is defined as the ratio of the percentage increase in DUV dose to the corresponding percentage decrease in total ozone. Figure 3 shows our calculated RAFs on the two-dimensional grid of latitude and season.

Considering the estimated ozone depletion in the range 1–5% through 1992 due to continued use of CFCs, as calculated by Pyle and Derwent with a two-dimensional model (Fig. 1 in ref. 6), we obtain the contour plots for the RAF of Fig. 3 when the erythral and DNA action spectrum is used. If we consider a global ozone depletion of 10% as an estimate of a steady state condition, we obtain RAFs between 2.1 and 2.5 for erythema and between 2.8 and 3.0 for DNA. In both cases the RAFs are larger than for the lower 1992 depletion estimates which indicates the non-linearity of the radiation amplification. Note from Fig. 2 that the ratio of increase in erythemally weighted UV-B dose to the decrease in total ozone amount is always greater than unity and quite close to 2 even in the equatorial region; this is in contrast to the results reported in ref. 6. Calculations by F. M. Luther (personal communication) performed after the present analyses and extensive verifications of our computational methods and data¹⁶ indicate that the differences between our results for RAF and those of Pyle and Derwent⁶ may be due to the different atmospheric data bases and different erythral action spectra used in the radiative transfer calculations. The uncertainty in our results introduced by the computational method is estimated to be at most 10% in RAF for all latitudes and seasons. In addition, the radiation amplification factor for DNA-weighted DUV is close to 3 and thus significantly larger than for erythemally-weighted DUV for all latitudes and seasons. As regards the damaging effects to the human population of increased UV radiation due to continued release of CFCs, the above results may be interpreted as follows. If a 5% global depletion of ozone is reached due to the continued release of CFCs, then an amplification of biologically

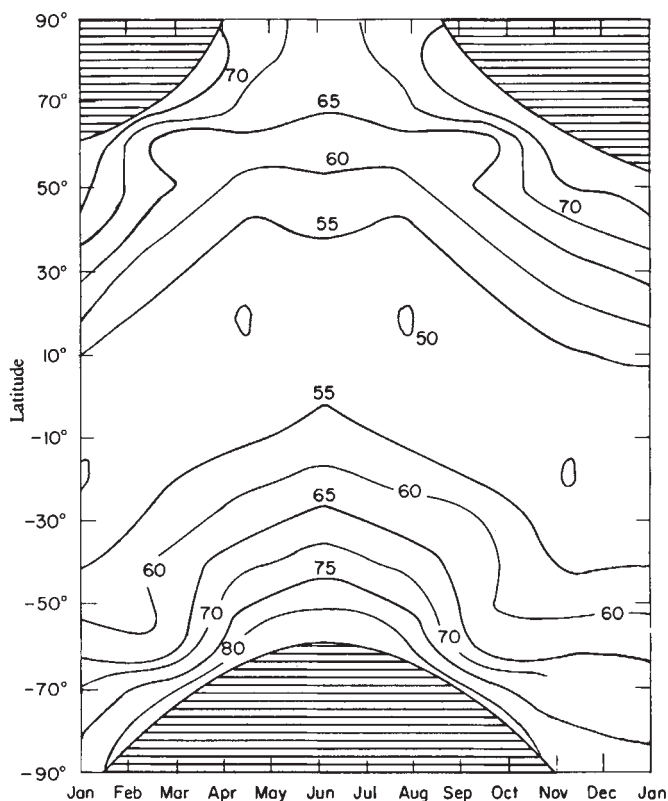


Fig. 2 Effective solar zenith angle determined as a daily average over continuously varying zenith angles at a given date and latitude. Shaded areas indicate polar regions for which no $\langle \mu_0 \rangle$ is determined.

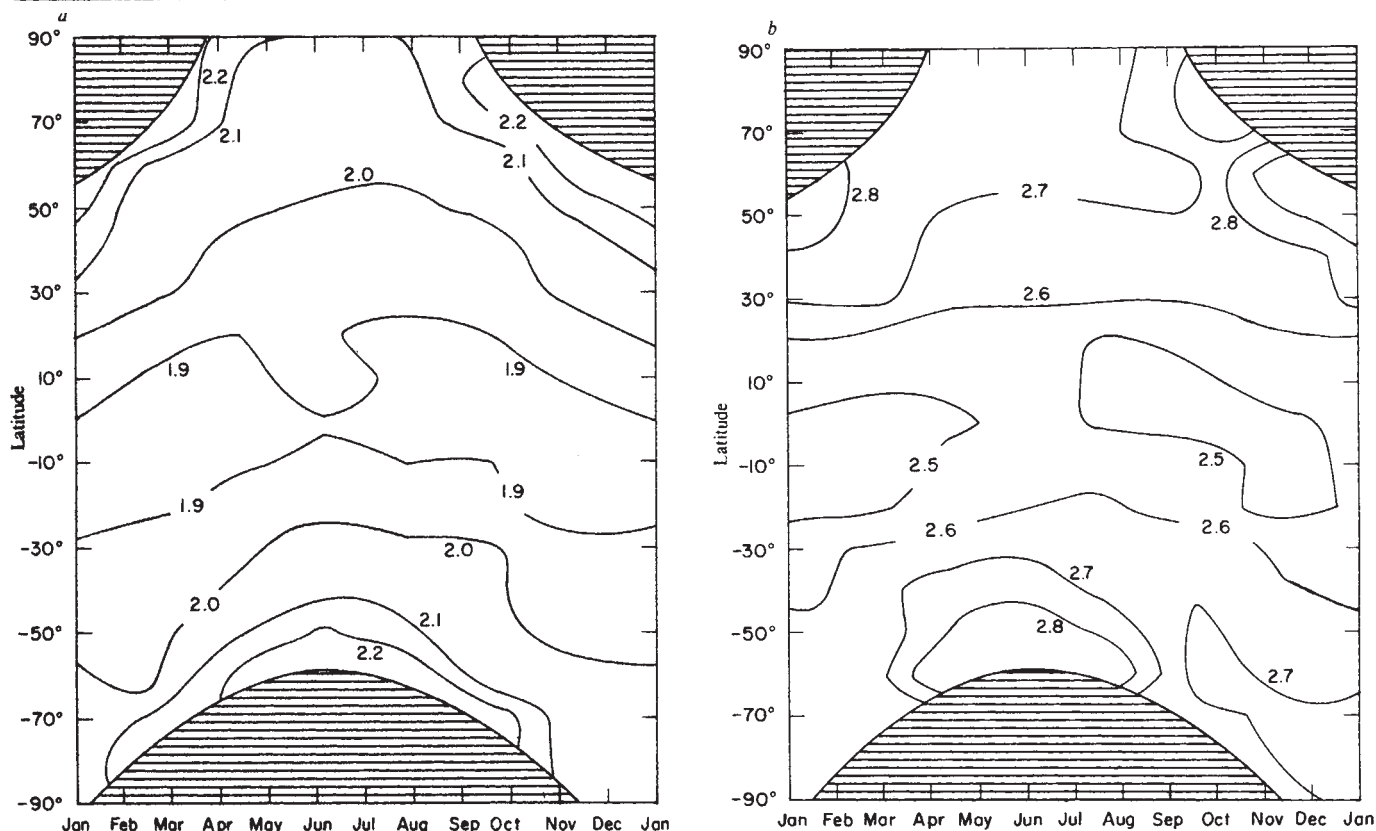


Fig. 3 Radiation amplification factor for *a*, erythemally weighted and *b*, DNA weighted UV-B dose resulting from continued CFC usage through 1992. Percentage depletion in the total ozone taken after ref. 6.

damaging radiation should be expected that increases erythema by ~10% (RAF ≈ 2.0) and DNA damage by about 13% (RAF ≈ 2.6). The estimated increase of skin cancer rates is likely to be of the same order.

In conclusion, as our modelling indicates, the approximate value of 2 for the RAF should be replaced by a non-linear functional relationship; the larger the ozone depletion the larger the RAF. Plotted as a function of latitude and season, however, the RAF shows a fairly small gradient in all directions, consistent with the results of Green and Mo⁷. Using the latitude- and season-dependent ozone depletions estimated by Pyle and Derwent for continued CFC usage to the year 1992, we arrive at an RAF range between 2.5 and 2.8 for DNA-weighted UV-B dose and between 1.9 and 2.2 for erythemally-weighted dose. (These new results have been confirmed by I. Isaksen *et al.*, unpublished data.) Note especially that our minimum RAF values are 1.9 (erythema) and 2.5 (DNA), which occur in the equatorial region.

We thank Dr E. P. Shettle for helpful discussions of atmosphere models and Dr J. Chang for making available the results of ref. 4.

Received 23 April; accepted 16 September 1981.

1. Molina, M. J. & Rowland, F. S. *Nature* **249**, 810–812 (1974).
2. *Protection against Depletion of Stratospheric Ozone by Chlorofluorocarbons* (National Academy of Sciences, Washington DC, 1979).
3. *Stratospheric Ozone Depletion by Halocarbons: Chemistry and Transport* (National Academy of Sciences, Washington DC, 1979).
4. Wine, P. H. *et al. J. geophys. Res.* **86**, 1105–1112 (1981).
5. Pyle, A. J. *Nature* **271**, 42–43 (1978).
6. Pyle, J. A. & Derwent, R. G. *Nature* **286**, 373–375 (1980).
7. Green, A. E. S. & Mo, T. *Impacts of Climatic Change on the Biosphere*, Chap. 2, Part 1, (US Dept of Transportation, CIAP Monogr. 5, Rep. DOT-TST-75-55, 1975).
8. Gerstl, S. A. W. *Application of the Adjoint Method in Atmospheric Radiative Transfer Calculations* Los Alamos National Laboratory Report LA-UR-80-17; *Application of Modern Neutron Transport Methods to Atmospheric Radiative Transfer*, Int. Rad. Symp., Fort Collins (1980).
9. McClatchey, R. A., Fenn, R. W., Selby, J. E. A., Volz, F. E. & Garing, J. S. *Optical Properties of the Atmosphere* 3rd edn (AFGL-72-0497, 1972).
10. Kneizys, F. X. *et al. Atmospheric Transmittance/Radiance: Computer Code LOWTRAN 5* (AFGL-TR-80-0067, 1980).
11. Shettle, E. P. & Green, A. E. S. *Appl. Opt.* **13**, 1567–1581 (1974).
12. Shettle, E. P. & Fenn, R. W. *Models for the Aerosols of the Lower Atmosphere* (AFGL-TR-79-0214, 1979).
13. Coblenz, W. W. & Stair, R. *U.S. Bur. Standard. J. Res.* **12**, 13–14 (1934).

14. Green, A. E. S. & Miller, J. H. in *Impacts of Climatic Change on the Biosphere*, 2–60 (US Dept of Transportation CIAP Monogr. 5, Rep. DOT-TST-75-55, 1975).
15. Setlow, R. B. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3363–3366 (1974).
16. Zardecki, A. & Gerstl, S. A. W. *Calculations of Solar Irradiances in Clear and Polluted Atmospheres and Potential Effects on Plant Life* (Los Alamos National Laboratory Rep. LA-9010-MS, 1981).

High species diversity and abundance of the epibenthic community in an oxygen-deficient basin

Verena Tunnicliffe

Ocean Ecology, Institute of Ocean Sciences, PO Box 6000, Sidney, British Columbia, Canada

Oxygen-deficient basins are common features of the ocean¹. Studies of the response of benthic communities to low-oxygen conditions have previously been restricted to sediment infauna, and have reported mass mortalities during sudden oxygen depletion^{2,4} and low species diversity during sustained oxygen deficiencies^{5,6}. The deep epibenthic fauna of a 220-m deep fjord in British Columbia experiences annual conditions of low oxygen (<1.0 ml l⁻¹). I have observed the reactions of the attached fauna to the changing oxygen levels during numerous dives in the submersible *Pisces IV* over 10 months, and report here that, despite records of the decimation of benthic invertebrate populations^{2,5,7–9}, this deep-water assemblage maintains a higher diversity and animal abundance than does much of the upper photic zone. The tolerance developed by this community for a wide range of oxygen conditions, including periods of anoxia, has interesting implications for physiology, behaviour and community evolution.

The glacier-carved fjords of British Columbia are narrow and deep and frequently have restricted water circulation; the vertical sides provide sediment-free surfaces for the settlement of

epibenthic organisms. Saanich Inlet, on Vancouver Island, is such a fjord, and has a maximum depth of 220 m. For much of the year the bottom waters are anoxic and anaerobic respiration is associated with the build-up of hydrogen sulphide^{1,11}. The oxygen deficiency is moderated by an annual flushing which brings dense, oxygenated water over the shallow sill at the mouth from the nearby Strait of Georgia^{12,13}.

Using the manned submersible *Pisces IV* (Institute of Ocean Sciences, BC), it was possible to study *in situ* the effects of an unstable anoxic layer on a hard substratum community in this fjord. Thirty-three dives were made from May 1980 to March 1981 throughout the Saanich Inlet. At the major study site (Fig. 1) the submersible ascended close to the cliff from 215 to 30 m. Water samples were collected *in situ* within 2 m of the cliff by a sampling system on the submersible and later analysed for dissolved oxygen. By extending the sampling tube, oxygen values at 0.05 m and 2 m off the rock face could be compared; no significant differences were found.

Photographs of the cliff were taken at 5-m intervals in depth using an external strobe light and an internal camera placed 1 m from a 30 × 30-cm quadrat extended from the submersible to touch the rock. The distributions and abundances of animals larger than 1 cm were determined for each photograph and the diversity calculated from the combined results. Species identifications were made from close-up photographs and verified from collections of loose rocks.

The dissolved oxygen structure for the sampling period is plotted in Fig. 2. Until late August 1980, the bottom 55 m of Saanich Inlet were anoxic. Dense, oxygenated water downwelled into the bottom layers and pushed a wedge of anoxic water upwards throughout September. In late October, the epibenthic fauna as shallow as 85 m was exposed to water with <0.1 ml O₂ per l. Re-oxygenation ceased after November and by May 1981 the bottom waters developed patches of anoxia.

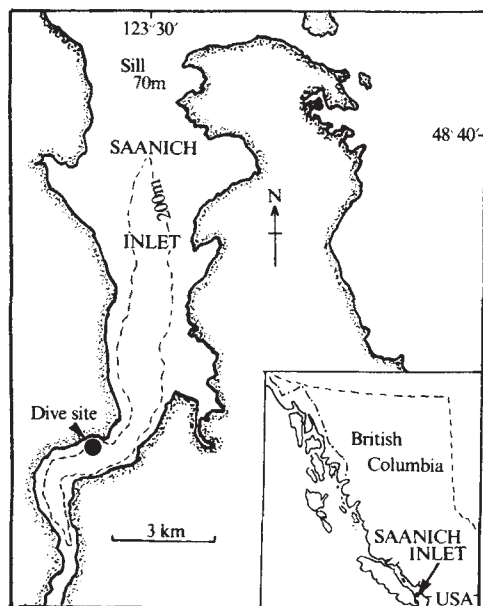


Fig. 1 Location of submersible dive site in Saanich Inlet, British Columbia. Deep-water circulation is restricted by the sill at the mouth of the inlet. In the south the walls are nearly vertical, exposing a silt-free gneissic substratum. Typical summer salinities range from 28.5‰ at the surface to 31.2‰ at 200 m; freshwater run-off into the inlet is minimal, the greatest influence being from the Cowichan River north of the sill from which intrusions into the inlet can be measured¹³. Suspended sediment load is correspondingly low, although dense turbid layers due to plankton may form¹³. Tidally influenced currents result in shallow-water exchange over the sill¹⁷ and these daily fluctuations are detectable in the middle of the inlet but at velocities that rarely exceed 10 cm s⁻¹ (data not shown). The euphotic zone, defined by light levels >1% of that at the surface, was <20 m in the spring.

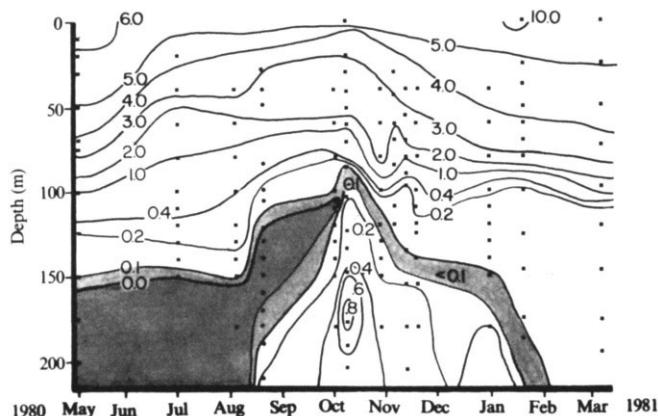


Fig. 2 Dissolved oxygen contours in ml O₂ per l for Squally Reach, Saanich Inlet, British Columbia. Sample points represent water drawn both by surface casts and by *Pisces IV* submersible; in the latter case, samples were fixed immediately inside the submersible and later analysed by the modified Winkler technique. Shading indicates anoxic zone. The disruption of contours from September to November 1980 is due to deep-water renewal by dense, oxygenated water entering the inlet over the sill.

Photographs and examination of samples in May 1980 showed the rock from the bottom to 130 m to be bare of macroscopic life. Small white sponges (*Suberites simplex*) appeared at 130 m and galatheid crabs (*Munida quadrispina*) at 127 m. In October, the anoxic layer was located between 100 and 130 m (Fig. 2). Tube worms were contracted, brachiopods closed and anemones limp. Dense populations of crab were found at 85 m (B. Burd, personal communication) where they had retreated as the anoxic water rose. By November brachiopod shells and crab carapaces were seen on ledges below 100 m; a few encrusting sponges were disintegrating and fragments of ascidian tunics were present between 85 and 100 m. Despite these observations, comparison of photographs taken above 120 m before and after the anoxic rise detected no significant changes in abundance of non-motile fauna. However, no epifauna were observable from the submersible between 130 and 120 m.

Two peaks in animal abundance (Fig. 3) were observed at 100 m and 75 m. At the lower depth, 84% of the individuals were sponges; at 75 m, 75% were ascideans. Numbers steadily declined towards the top station at 30 m and cover was correspondingly low. Above 30 m a distinct algae-dominated shallow-water fauna was present.

Diversity (Fig. 3) between 85 and 100 m was surprisingly great in view of the prevailing low-oxygen conditions. At 60 m it

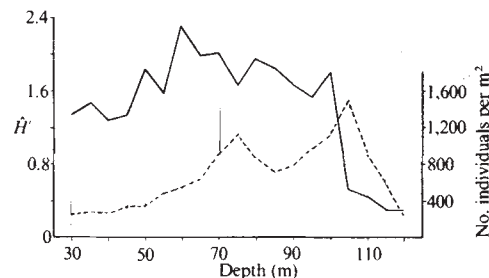


Fig. 3 Faunal diversity (—) and animal abundance (---) as measured from photographs from the submersible.

$$\hat{H}' = -\sum_{i=1}^n p_i \ln p_i$$

(Shannon-Weaver diversity index)¹⁸. Four photographs were used for each depth from submersible ascents on the same location in September and November 1980, and February and late March 1981; qualitative comparisons were made with photographs taken without the quadrat in May 1980. Pictures were taken only where the rock was within 15° of vertical and the submersible could manoeuvre close to the rock. The resulting photographs of the 30 × 30-cm quadrat were readily reproduced to life size and so analysed for animals larger than 1 cm.

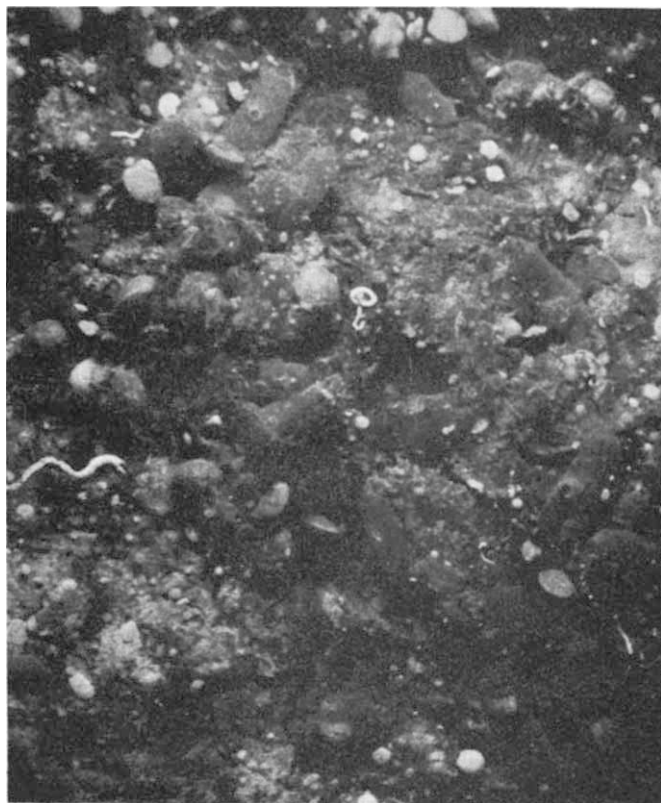


Fig. 4 Photograph from *Pisces IV* at -85 m. Here, the epifauna is dominated by ascideans but brachiopods, sponges and hydroids can also be seen.

marked the point at which animals of both shallow-water and high-oxygen dependence overlapped with the deeper, more tolerant species. In the photic zone diversity dropped with the appearance of encrusting red algae at 50 m. More than 40 species of epibenthic animals were identifiable from the submersible between 120 and 85 m and many more from rock samples. Phyletic diversity was high in this zone; 10 phyla were represented, including 16 classes. Conversely, infaunal assemblages of oxygen-poor areas were mostly impoverished⁵⁻⁷.

The oxygen levels at which the greatest faunistic changes are reported to occur are 1.5 and 1.0 ml l⁻¹ (refs 6, 9). Despite <1.0 ml l⁻¹ of oxygen for half the monitored time, the zone between 85 and 100 m had neither low diversity nor small populations (Fig. 4); even shelled animals such as brachiopods were common, despite possible physiological dissolution of carbonate⁶.

The flushing of Saanich Inlet and the rise of anoxic water is an annual event^{1,3,14}; 0.5 ml O₂ per l has been recorded¹ as shallow as 60 m. Elsewhere, similar influxes of low-oxygen water have resulted in the demise of large areas of benthic fauna^{3,4,7,8}. Either a benthic community repeatedly re-establishes itself after oxygen deficiency^{5,7} or it becomes highly tolerant of changing oxygen conditions. The latter seems to be the case in Saanich Inlet; the community maintained a stable constituency, noticeable mortality occurred only in areas of sustained anoxia and throughout the 10 months, there was no significant recruitment to any taxon.

The persistence of this community is pertinent to (1) the behavioural and physiological adaptations that allow survival in very low-oxygen conditions; (2) the possible regulation of competitive dominance^{10,15} in the substrate-limited depths by periods of anoxia that cause some mortality; and (3) the evolution of community structure during the Palaeozoic when atmospheric levels of oxygen were low¹⁶—the composition and behaviour of the deeper part of this community may reflect patterns in early Cambrian assemblages.

Without the continued support of the crew of *Pisces IV*, this work would not be possible. I thank B. Burd, Dr R. Brinkhurst and, especially, G. Calderwood.

Received 15 June; accepted 29 September 1981.

1. Richards, F. A. in *Chemical Oceanography* Vol. 1 (eds Riley, J. P. & Skirrow, G.) 611–645 (Academic, London, 1965).
2. Brongersma-Sanders, M. *Mem. geol. Soc. Am.* **67**, 941–1008 (1957).
3. Levings, C. D. in *Fjord Oceanography* (eds Freeland, H. J. et al.) 515–522 (Plenum, New York, 1979).
4. Tulkki, P. *Cah. Biol. mar.* **6**, 455–463 (1965).
5. Rosenberg, R. *J. exp. mar. Biol. Ecol.* **26**, 107–133 (1977).
6. Rhoads, D. C. & Morse, J. W. *Lethaea* **4**, 413–428 (1971).
7. Leppakoski, E. *Cah. Biol. mar.* **10**, 163–172 (1969).
8. Kitching, J. A. et al. *J. Anim. Ecol.* **45**, 731–758 (1976).
9. Rosenberg, R. in *Fjord Oceanography* (eds Freeland H. J. et al.) 499–514 (Plenum, New York, 1979).
10. Connell, J. H. *Science* **199**, 1302–1310 (1978).
11. Anderson, J. J. & Devol, A. H. *Estuar. coastal mar. Sci.* **1**, 1–10 (1973).
12. Gade, H. G. & Edwards, A. in *Fjord Oceanography* (eds Freeland, H. J. et al.) 453–490 (Plenum, New York, 1979).
13. Herlinveaux, R. H. *J. Fish. Res. Bd Can.* **19**, 1–37 (1962).
14. *Pac. mar. Rep. Ser.* 80-5 (Inst. Ocean Sciences, Sidney, British Columbia, 1980).
15. Levin, S. & Paine, R. T. *Proc. natn. Acad. Sci. U.S.A.* **71**, 2744–2747 (1974).
16. Berkner, L. V. & Marshall, L. C. *J. atmos. Sci.* **22**, 225–261 (1965).
17. Herlinveaux, R. H. *Fish. Res. Bd Can. tech. Rep.* **99**, (1968).
18. Pielou, E. C. *An Introduction to Mathematical Ecology* (Wiley, New York, 1969).

Epithelial cells of *Hydra* are dye-coupled

Scott E. Fraser* & Hans R. Bode†

* Department of Physiology and Biophysics and † Department of Developmental and Cell Biology and the Developmental Biology Center, University of California, Irvine, California 92717, USA

In the past decade, a strong correlation has been established between gap junctions, seen in cell ultrastructure studies, and cell coupling (ionic, metabolic or dye coupling) assayed physiologically¹. In *Hydra*, ultrastructural analyses have indicated that the epithelial cells of both cell layers are connected extensively by gap junctions; gap junctions have also been observed between the two layers²⁻⁷. On the basis of these results, one would expect electrical and dye coupling between epithelial cells of *Hydra*. However, de Laat et al.⁸ reported that these cells were neither dye- nor electrically coupled, which was unexpected as cells in another coelenterate have been shown to be coupled ionically⁹. Cell-cell coupling in *Hydra* is particularly interesting because extensive experiments on head regeneration in this coelenterate have led to well-defined models of patterning that require communication between cells of the type that may be provided by gap junctions¹⁰⁻¹⁴. We have re-examined dye coupling in *Hydra* and we report here that, after injection of Lucifer yellow into single epithelial cells, neighbouring cells were observed to contain the dye. We conclude that the epithelial cells of *Hydra* are indeed dye-coupled.

The body column of *Hydra* consists of two layers of epithelial cells, the epidermis and the gastrodermis, arranged in a cylindrical shell with the epidermis on the outside and the gastrodermis lining the gastric cavity. The cells of both layers possess large vacuoles; results of light and electron microscopy studies suggest that these vacuoles occupy less than half the cell volume^{15,16}. By fixing the animals in isotonic solutions instead of the typical hypertonic solutions, it has been shown (R. D. Campbell, personal communication) that the vacuole actually comprises a much greater fraction (perhaps 80–90%) of the cell volume. The cytoplasm is therefore often only a thin shell 1–3 µm thick between the plasma membrane and the vacuole—electrophysiological properties support this arrangement. On penetrating an ectodermal epithelial cell with a glass micro-electrode, several investigators have found a transient negative membrane potential, followed by a stable positive potential¹⁷⁻¹⁹. Chain¹⁹ reported that the electrode could then be moved large distances into the cell without altering the positive potential, consistent with the electrode being within the vacuole. Movement of the animal has also hindered intracellular elec-

trophysiological investigations of *Hydra*. During impalement, movement of only a few cell diameters is sufficient to interfere with measurements. Impaling an otherwise quiet *Hydra* can induce movement, and use of nerve-free *Hydra*^{20,21}, which are not moving on a macroscopic scale, is of little help as they still make small movements. Thus the delicate process of penetrating and remaining in the thin shell of cytoplasm is made even more difficult.

To increase the likelihood of positioning the electrode in the cytoplasm of the cell, we used the following techniques. Movement of *Hydra* was eliminated by: (1) Removing the head and foot of the animal, which greatly reduced stretching and contraction of the body column. (2) The body column was threaded onto monofilament fishing line having a diameter slightly larger than that of the body column. The column easily stretched circumferentially to accommodate the larger diameter. (This is a standard technique used in *Hydra* grafting experiments^{13,14,22,23}.) (3) Animals were starved for 2–4 days (*Hydra* can starve for 10–15 days and still remain healthy). To reduce vacuole size, animals were acclimatized to 50 mM sucrose in *Hydra* medium for 18–24 h, which does not affect the health of the animal¹⁸. With these treatments, stable negative or positive membrane potentials were readily obtained.

Epithelial cells were impaled with micropipettes containing Lucifer yellow and filled with the dye iontophoretically. While developing methods to optimize cell impalement, we observed that the cells are coupled in a wide variety of conditions (normal, sucrose and hypertonic medium), which indicates that coupling of the cells is not an abnormal state induced by our manipulation of the conditions. Of the 121 cells that were impaled, 44 were located by fluorescence microscopy; for 19 of these, Lucifer yellow was found in at least one neighbouring cell. Presumably, a large number of cells could not be found due to cell death.

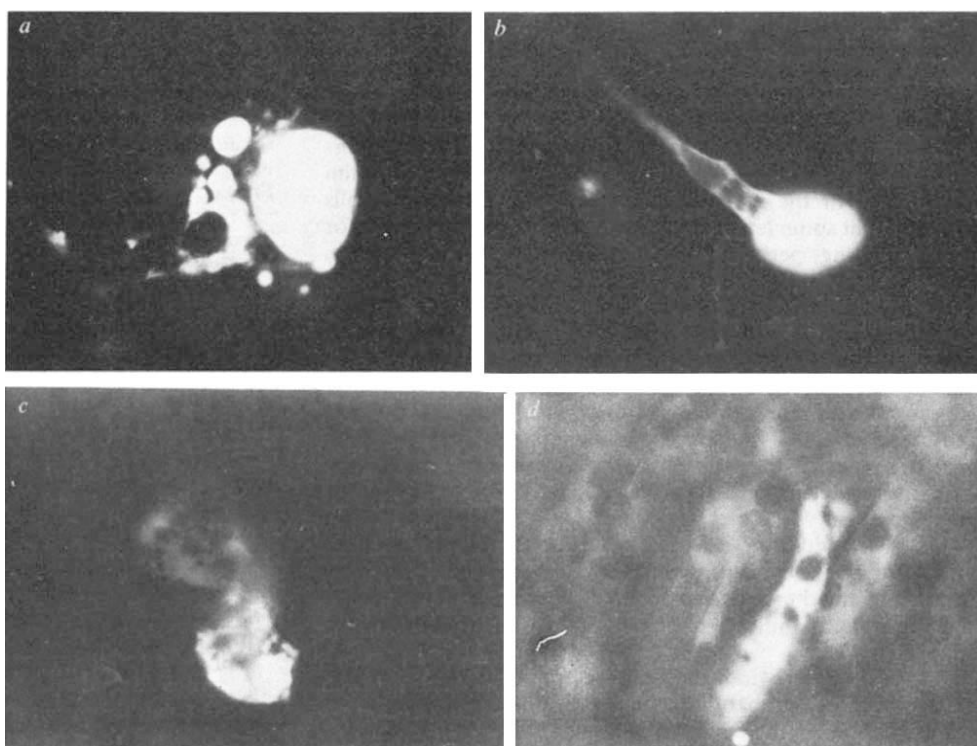
Occasionally, cells lysed while being observed in the fluorescence microscope. No residual dye remained after such lysis and the neighbouring cells did not pick up any dye.

In subsequent experiments using the improved methods described above, 86 cells were impaled and filled with Lucifer yellow: 45 of these were located by fluorescence microscopy, and 25/45 cases demonstrated passage of the dye to one or more neighbouring cells. Four examples of dye-injected cells are shown in Fig. 1. In all cases, the filled cells were epithelial cells, as opposed to the much smaller cells found in the interstices between epithelial cells. They were of the same size and shape as epithelial cells and often had long processes extending away from the cell, which were probably muscle processes found at the base of each epithelial cell type²⁴ (see Fig. 1b).

Two distinct patterns of dye distribution were observed. In one, the cytoplasm was uniformly, intensely stained, while several round dark spots (assumed to be vacuoles and the nucleus) were not (Fig. 1d). In the other pattern, one of the vacuoles was very intensely stained while the remainder were not. In these cells, the cytoplasm was much less intensely stained (Fig. 1a, b). Although no dye passage was seen in these cases (Fig. 1a, b), dye coupling was observed for both types of filled cell (Fig. 1c, d). The vacuoles of the cells coupled to the injected cell were always dark compared with the cytoplasm. Coupling was observed both within each cell layer and between cell layers.

Two types of membrane potential were also found. In some cases, the membrane potential was negative -10 ± 6 mV ($\pm$ s.d.) before, during and after filling the cell with Lucifer yellow. In other cases, a positive membrane potential was obtained ($+21 \pm 8$ mV) before, during and after dye filling. Occasionally, the initial potential was negative, and suddenly switched to positive during dye filling. The reverse switch was never observed. These potentials are probably not an exact measure of the

Fig. 1 Sample micrographs of *Hydra* cells filled with Lucifer yellow. *a*, *b*, A region with positive membrane potential was filled with the dye and no passage occurred. In *a*, there was no leakage of dye from the vacuoles; in *b*, the cytoplasm was clearly labelled as well as the single bright vacuole. Presence of Lucifer yellow in the cytoplasm was always easily recognized by the fine cell processes that were filled with the dye. The vacuole in *a* fragmented into many smaller vacuoles during observation, before the photograph was taken. *c*, Dye passage in a region of positive membrane potential which was dye filled. A bright apical vacuole was observed in the cell, with considerable amounts of dye in the surrounding cytoplasm. The dye passed to two cells, one of which is in the plane of focus. *d*, A region of negative membrane potential that was dye filled. The vacuoles in all these cells were dark. The dye passed to many other cells, extending beyond the limits of the photograph. The cell shape observed in *a*–*d* varies because the shape of the animals at the time of observation. Cells were identified (epidermis as opposed to gastrodermis) by their plane of focus with respect to the mounted stenoteles, which are known to be on the outer surface of *Hydra*³⁰. *Hydra attenuata* were maintained in *Hydra* medium (1 mM CaCl₂, 1 mM NaHCO₃, 10 μ M EDTA), and were starved for 2–4 days before an experiment. For the 18–24 h preceding an experiment, the medium was supplemented with 50 mM sucrose (to reduce vacuole size). The head and basal disk of the animal were removed and the resulting body column threaded on to a monofilament fishing line (0.25 mm diameter). The fishing line was secured with dental wax in the centre of a Petri dish filled with sucrose–*Hydra* medium (with body column parallel to the dish bottom). The dish was mounted on the stage of a Hoffman modulation microscope (viewed at $\times 280$); illumination was supplied by a fibre optic light source placed outside the Faraday cage, with heat filters and a red filter positioned in the light path to avoid damaging the Lucifer yellow-filled cells. Glass micropipettes (200–800 Ω) were pulled from Ultratip glass (Haer) and black-filled with 3% Lucifer yellow CH (Aldrich). Cells were impaled using a piezoelectric bending element which withdrew the electrode slightly (2 μ m) and then allowed it to snap forward, thus penetrating the cell membrane. Lucifer yellow was introduced into the cell iontophoretically with 0.5 nA hyperpolarizing current pulses (1 Hz, 50% duty cycle) for 15 s. The animals were removed from the fishing line and allowed to recover in the dark for 10–60 min, after which time they were pressed between two microscope coverglasses in a small amount of sucrose–*Hydra* medium. A small bead of silicone grease (Dow-Corning) placed between the two coverglasses prevented the animals from being overly flattened. This 'sandwich' arrangement could be flipped to view both sides of the animal. The animals were observed and photographed ($\times 325$) with a Zeiss UEM epifluorescence microscope through a Zeiss filter set designed for Lucifer fluorescence (487704).



The head and basal disk of the animal were removed and the resulting body column threaded on to a monofilament fishing line (0.25 mm diameter). The fishing line was secured with dental wax in the centre of a Petri dish filled with sucrose–*Hydra* medium (with body column parallel to the dish bottom). The dish was mounted on the stage of a Hoffman modulation microscope (viewed at $\times 280$); illumination was supplied by a fibre optic light source placed outside the Faraday cage, with heat filters and a red filter positioned in the light path to avoid damaging the Lucifer yellow-filled cells. Glass micropipettes (200–800 Ω) were pulled from Ultratip glass (Haer) and black-filled with 3% Lucifer yellow CH (Aldrich). Cells were impaled using a piezoelectric bending element which withdrew the electrode slightly (2 μ m) and then allowed it to snap forward, thus penetrating the cell membrane. Lucifer yellow was introduced into the cell iontophoretically with 0.5 nA hyperpolarizing current pulses (1 Hz, 50% duty cycle) for 15 s. The animals were removed from the fishing line and allowed to recover in the dark for 10–60 min, after which time they were pressed between two microscope coverglasses in a small amount of sucrose–*Hydra* medium. A small bead of silicone grease (Dow-Corning) placed between the two coverglasses prevented the animals from being overly flattened. This 'sandwich' arrangement could be flipped to view both sides of the animal. The animals were observed and photographed ($\times 325$) with a Zeiss UEM epifluorescence microscope through a Zeiss filter set designed for Lucifer fluorescence (487704).

Table 1 Correlation between membrane potentials and staining patterns of *Hydra* epithelial cells

Staining pattern	Type of membrane	potential (initial/final)	
	-/- (n = 38)	-/+ (n = 10)	+/+ (n = 38)
Cells that could be located	7	9	29
a, Bright cytoplasm and dark vacuoles	7	2	5
No. of cells coupled	4	1	3
b, Bright vacuole and dim cytoplasm	0	7	21
No. of cells coupled	-	4	10
c, Ambiguous	0	0	3
No. of cells coupled	-	-	2

Data from the final set of animals in which the membrane potential was measured before, during and after the injection of Lucifer yellow. The cells were grouped together by the membrane potentials before and after injection. Techniques used are described in Fig. 1 legend. The cytoplasm of all five samples with positive, positive potentials (+/+) and one with negative, positive potentials (-/+) in *a* were much less intensely stained. This low level of observed dye could be due to dye passage from a filled cell which subsequently lysed, accounting for the cytoplasmic dye location. In three of these six cases, fluorescent debris was observed around the dimly fluorescent cells.

membrane potential because the size of the electrode tip potential is unknown. A contribution from the tip potential could be substantial due to the large difference between the ionic strength of *Hydra* medium (5 mM) and that of the cytoplasm (120–150 mM) (ref. 25 and H.R.B., unpublished results). Nevertheless, two distinct cell regions of differing membrane potential were consistently observed throughout our experiments.

A correlation was observed between the staining patterns and membrane potentials (see Table 1): intensely stained vacuoles and dimly labelled cytoplasm were correlated with positive potentials. These results are consistent with the positive potential resulting from penetrating a vacuole. The appearance of dim fluorescence in the cytoplasm of the positive potential fills indicated that some leakage of the dye occurred during and/or after dye filling, perhaps when the electrode was withdrawn. Although there were equal numbers of cells filled with dye for each type of membrane potential, a much smaller fraction (18% as opposed to 76%) of the 'negative, negative' fills were later located, suggesting a higher probability of cell damage when impaling the cytoplasm.

It seems likely that in previous work⁸, which recorded positive membrane potentials for many of the cells impaled, the electrode was placed in the vacuole, as we found a strong correlation between positive potentials and dye filling of the vacuole. This may have greatly reduced the chance of observing ionic or dye coupling. In addition, cell damage or cell death may have uncoupled the cells. Approximately half the cells that were successfully filled with dye in our work showed no dye passage, which we take to indicate the ease with which the cells may be damaged.

Our results clearly indicate that the epithelial cells of *Hydra* are dye-coupled, probably through gap junctions, as has been well established in the epithelia of other animals. This finding is of particular interest in *Hydra* in terms of pattern formation. Two developmental gradients, one of activation and one of inhibition, are involved in the formation of the head at the apical end^{8,10,14,26}, and a further two gradients control foot formation^{10,13,14}. Indirect evidence suggests that the inhibition gradients are due to diffusible substances^{13,27,28}. Recently, it has been shown that the epithelial cells are responsible for transmission of the head inhibition down the body column²⁹. A logical channel for the movement of such a diffusible substance would be through the gap junctions of the epithelial cells.

We thank Dr M. Bronner-Fraser and P. Bode for critical reading of the manuscript. This work was supported by grants Hd08086 and GM29130 and NSF grant BNS-8023638.

Received 31 July; accepted 29 September 1981.

1. Gilula, N. B. in *International Cell Biology* (eds Brinkley, B. R. & Porter, K. R.) 61–69 (Rockefeller University Press, New York, 1977).
2. Hand, A. R. & Gobel, S. J. *J. Cell Biol.* **52**, 397–408 (1972).
3. Filshie, B. K. & Flower, N. E. *J. Cell Sci.* **2**, 151–172 (1977).
4. Wood, R. L. *J. ultrastruct. Res.* **58**, 299–315 (1977).
5. West, D. L. *Tissue Cell* **10**, 629–646 (1978).
6. Wood, R. L. *Cell Tissue Res.* **199**, 319–388 (1979).
7. Lepault, J., McDowall, A. W. & Grimmelikhuijzen, C. J. P. *Cell Tissue Res.* **209**, 217–224 (1980).
8. deLaat, S. W., Tertoolen, L. G. J. & Grimmelikhuijzen, C. J. P. *Nature* **288**, 711–713 (1980).
9. Josephson, R. K. & Schwab, W. E. *J. gen. Physiol.* **74**, 213–236 (1979).
10. Wolpert, L. *Curr. Topics devl Biol.* **6**, 183–224 (1971).
11. Gierer, A. & Meinhardt, H. *Kybernetik* **12**, 30–39 (1972).
12. Meinhardt, H. & Gierer, A. *J. Cell Sci.* **15**, 321–346 (1974).
13. MacWilliams, H. K. *Devl Biol.* (in the press).
14. MacWilliams, H. K. *Devl Biol.* (in the press).
15. Wanek, N., Marcum, B. A. & Campbell, R. D. *J. exp. Zool.* **212**, 1–12 (1980).
16. Graf, L. & Gierer, A. *Wilhelm Roux Arch. devl Biol.* **188**, 141–151 (1980).
17. Macklin, M. J. *cell. Physiol.* **70**, 191–196 (1967).
18. Benos, D., Kirk, R. G., Barba, W. P. & Goldner, M. M. *Tissue Cell* **9**, 11–22 (1977).
19. Chain, B. M. *J. exp. Biol.* **88**, 161–173 (1980).
20. Campbell, R. D. *J. Cell Sci.* **21**, 1–13 (1976).
21. Marcum, B. A. & Campbell, R. D. *J. Cell Sci.* **29**, 17–33 (1978).
22. MacWilliams, H. K. & Kaftos, F. C. *Science* **159**, 1246–1248 (1968).
23. Rubin, D. I. & Bode, H. R. *Devl Biol.* (in the press).
24. Hess, A., Cohen, A. & Robson, E. Q. *Jl microsc.* **98**, 315–326 (1957).
25. Benos, D. J. & Prusch, R. D. *Comp. Biochem. Physiol.* **43A**, 165–171 (1972).
26. Webster, G. *Biol. Rev.* **46**, 1–46 (1971).
27. Wolpert, L., Clarke, M. R. B. & Hornbush, A. *Nature new Biol.* **239**, 101–105 (1972).
28. MacWilliams, H. N. *Am. Zool.* **14**, 633–645 (1974).
29. Rubin, D. I. & Bode, H. R. *Devl Biol.* (in the press).
30. Bode, H. R. & Flick, K. M. *J. Cell Sci.* **21**, 15–34 (1976).

Protein phosphatase-1 is involved in *Xenopus* oocyte maturation

Denise Huchon*, René Ozon*

& Jacques G. Demaille†‡

* Laboratoire de Physiologie de la Reproduction, Groupe Stéroïdes, ERA CNRS 964, Université Pierre et Marie Curie, 4 Place Jussieu, 75230 Paris Cédex 05, France

† Centre de Recherches de Biochimie Macromoléculaire du CNRS, Faculté de Médecine et U-249 INSERM Biochimie des Régulations des Systèmes Contractiles, BP 5015, 34033 Montpellier, France

Two key steps in meiotic maturation of the *Xenopus* oocyte involve protein phosphorylation–dephosphorylation¹. A burst of cyclic AMP-independent phosphorylation occurs at the time of the breakdown of the nuclear envelope, whether maturation is triggered by progesterone, by the inhibitor of cyclic AMP-dependent protein kinase (PKI)² or by the maturation promoting factor (MPF)^{1,3}. Also, an *in ovo* decrease in the level of the free catalytic (C) subunit of cyclic AMP-dependent protein kinase, induced by microinjection of pure PKI⁴, initiates meiotic maturation^{1,2,4}. This suggests that MPF appears only after dephosphorylation of a phosphorylated maturation protein Mp-P⁴. We now show, by microinjection of pure inhibitor-1 of protein phosphatase-1⁵, which blocks progesterone- and PKI-induced maturation but not the MPF-induced one, that protein phosphatase-1 catalyses the dephosphorylation of Mp-P.

Phosphoinhibitor-1 was injected into *Xenopus* oocytes to a final concentration of 15 μ M. Microinjected oocytes maintained in saline remained morphologically healthy and resumption of meiosis never occurred. When exposed to 1 μ M progesterone 4 h after microinjection of inhibitor-1, they matured with the same efficiency and kinetics as controls, showing that inhibitor-1 is neither an inducer of maturation by itself nor toxic to the cell. Oocytes were also stimulated by 1 μ M progesterone at various times before and after the microinjections (Fig. 1). Inhibitor-1 proved to be a potent inhibitor of maturation, having maximal efficacy when injected 1 h after incubation with the steroid. In contrast, microinjection of inhibitor-1 3 h after progesterone

‡ To whom correspondence should be addressed.

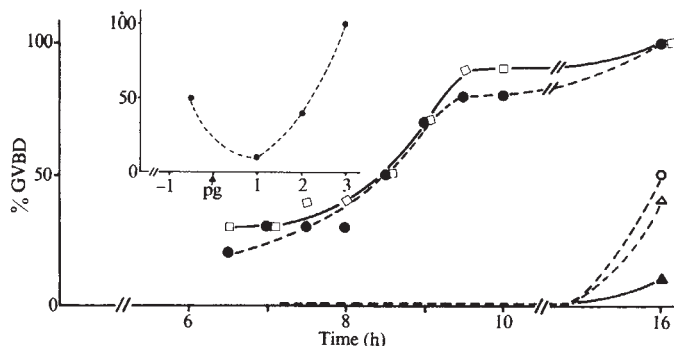


Fig. 1 Comparison of kinetics and % maturation after treatment with progesterone and phosphatase inhibitor-1. Ten oocytes were exposed either to 1 μ M progesterone only (●) or to 1 μ M progesterone 30 min after (○), 1 h before (▲), 2 h before (△) or 3 h before (□) inhibitor microinjection. Inset shows the time dependence of inhibitor efficacy: % of germinal vesicle breakdown (GVBD) 16 h after exposure to progesterone (pg) versus time of microinjection. Fully grown oocytes (stage VI₁₁) were defolliculated as described elsewhere⁴. Inhibitor-1 of phosphoprotein phosphatase-1 was purified according to Nimmo and Cohen⁵ and phosphorylated⁹ by the pure catalytic subunit of cyclic AMP-dependent protein kinase¹². Phosphoinhibitor-1 was freed from the catalytic subunit and MgATP by heat treatment and dialysis with 60 mM ammonium bicarbonate, lyophilized and dissolved in a 5 mM MES (2(*N*-morpholino) ethane sulphonate) buffer pH 7.0, containing 1 mg of serum albumin per ml. Microinjection at the equator level of 50 nl of a 100 μ M solution yielded an intracellular concentration of 15 μ M, assuming a diffusion compartment of 0.3 μ l (ref. 4). As inhibitor-1 was calculated to be 1.5 μ M in skeletal muscle⁵, microinjected inhibitor-1 should block all protein phosphatase-1 present. Its dephosphorylation into the inactive dephospho form¹⁰ is indeed prevented by the high free C subunit level of the resting oocyte⁴. The criterion for maturation was the appearance of a white spot at the animal pole of the oocyte. This experiment was done using oocytes from the same female; similar results were obtained with oocytes from two other females.

stimulation failed to alter the rate and extent of maturation, suggesting that inhibitor-1 blocks the steroid-induced maturation at an early step, without interfering with the effects of MPF. Indeed, microinjection of MPF into oocytes already injected with inhibitor-1 induced normal maturation (Fig. 2). The site of action of inhibitor-1 in the events triggered by progesterone was more precisely delineated by inducing maturation with pure PKI⁴. Inhibitor-1 inhibited the effect of subthreshold doses of PKI ($\leq 0.3 \mu$ M) while maturation induced by higher PKI concentration was only delayed (Fig. 3).

These findings confirm the four-step scheme of meiotic maturation proposed previously⁴: on reassociation of regulatory and catalytic subunits of cyclic AMP-dependent protein kinases, resulting from the progesterone-induced fall in the cyclic AMP levels, a phosphorylated maturation protein Mp-P is dephosphorylated to the active form Mp, which triggers the synthesis of

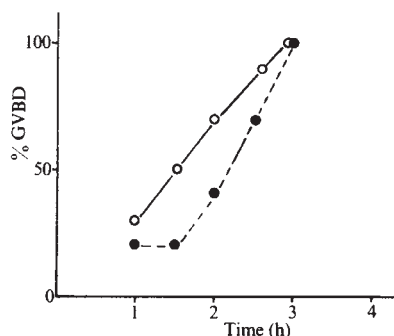


Fig. 2 Comparison of kinetics and % maturation after microinjections of phosphatase inhibitor-1 and MPF. Ten oocytes were injected either with MPF (●) or with inhibitor-1 and then immediately with MPF (○).

MPF. If this scheme were correct, progesterone- and PKI-induced maturation, but not the MPF-induced one, should be blocked by phosphatase inhibitors, as shown here to be the case. The efficacy of inhibitor-1 shows that the putative Mp-P is a substrate of its specific target, protein phosphatase-1, and that a type-2 phosphatase cannot be the major enzyme involved in Mp-P dephosphorylation⁶. This confirms previous reports of protein phosphatase-1 in a number of tissues and its catalysis of the dephosphorylation of various substrates belonging to different metabolic pathways⁷. It also eliminates initiation factor eIF-2 from this step in maturation because eIF-2 phosphatase was recently shown to be a type-2 phosphatase⁸. However, the remote possibility that inhibitor-1 blocks maturation by an unknown indirect mechanism cannot be excluded, even though protein phosphatase-1 is the only known target of inhibitor-1.

The fact that inhibitor-1 is more efficient when injected 1 h, rather than 2 h after the exposure to the steroid, indicates a slow, rate-limiting dephosphorylation of Mp-P into Mp, resulting from the balance between reduced protein kinase activity and a phosphatase activity that is either constant or enhanced by the dephosphorylation of endogenous inhibitor-1.

That oocytes escape at least in part the inhibitor-1-induced maturation blockade after a prolonged time is readily understood from the fact that inhibitor-1 does not inhibit its own

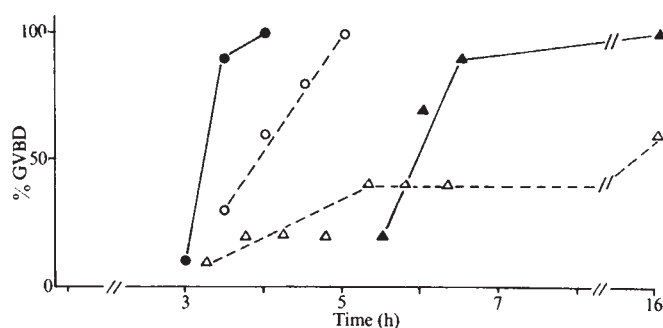


Fig. 3 Comparison of kinetics and % maturation after phosphatase inhibitor-1 and PKI treatment. PKI was microinjected 1 h after microinjection of inhibitor-1. Ten oocytes were injected either with 15 μ M inhibitor-1 and 1.5 μ M PKI (○) or with 15 μ M inhibitor-1 and 0.3 μ M PKI (△) or with 1.5 μ M PKI (●) or with 0.3 μ M PKI (▲). This experiment was done using oocytes from the same female; similar results were obtained with oocytes from two other females. PKI was prepared and injected as described previously^{4,13}.

dephosphorylation by protein phosphatase-1⁹. Microinjection of high concentrations of PKI only delays maturation because the total inhibition of the catalytic subunit of cyclic AMP-dependent protein kinase leads to rapid inactivation of inhibitor-1 by protein phosphatase-1 which dephosphorylates the inhibitor into its inactive dephospho form¹⁰.

This work also raises the intriguing possibility that inhibitor-1 is the maturation protein *per se*, although there is no experimental evidence for such identification.

The research was supported by DGRST, INSERM, CNRS and Fondation pour la Recherche Médicale.

Received 3 August; accepted 2 October 1981.

1. Maller, J. L. & Krebs, E. G. *Curr. Topics Cell Regulation* **16**, 271-311 (1980).
2. Maller, J. L. & Krebs, E. G. *J. biol. Chem.* **252**, 1712-1718 (1977).
3. Bellé, R., Boyer, J. & Ozon, R. *Gameete Res.* **2**, 137-145 (1979).
4. Huchon, D., Ozon, R., Fischer, E. H. & Demaille, J. G. *Molec. cell. Endocr.* **22**, 211-222 (1981).
5. Nimmo, G. A. & Cohen, P. *Eur. J. Biochem.* **87**, 341-351 (1978).
6. Ingebritsen, T. S., Foulkes, J. G. & Cohen, P. *FEBS Lett.* **119**, 9-15 (1980).
7. Burchell, A., Foulkes, J. G., Cohen, P. T. W., Condom, G. D. & Cohen, P. *FEBS Lett.* **92**, 68-72 (1978).
8. Stewart, A. A., Crouch, D., Cohen, P. & Safer, B. *FEBS Lett.* **119**, 16-19 (1980).
9. Nimmo, G. A. & Cohen, P. *Eur. J. Biochem.* **87**, 353-365 (1978).
10. Huang, F. L. & Glinsmann, W. H. *Eur. J. Biochem.* **70**, 419-426 (1976).
11. Dumont, J. N. *J. Morph.* **136**, 153-180 (1972).
12. Peters, K. A., Demaille, J. G. & Fischer, E. H. *Biochemistry* **16**, 5691-5697 (1977).
13. Demaille, J. G., Peters, K. A. & Fischer, E. H. *Biochemistry* **16**, 3080-3086 (1977).

Perivascular infiltrates of leukocytes in brains of scrapie-infected mice

C. Museteanu & H. Diringer

Robert Koch-Institut des Bundesgesundheitsamtes,
Nordufer 20, D-1000 Berlin 65, FRG

Four chronic subacute spongiform encephalopathies—Kuru and Creutzfeldt-Jakob diseases of man, transmissible encephalopathy in mink and scrapie in sheep—are caused by slow viruses with unconventional biological and biochemical properties¹⁻⁴. Since the transmission of the scrapie agent to mice⁵ and the comparative histological analysis of Kuru in man⁶ and of scrapie in the mouse, scrapie-mouse systems have become important models of spongiform encephalopathies. The predominant histopathological lesions (for review see refs 2, 4) found in these diseases result from nerve cell degeneration⁷. It seems that, in contrast to diseases of the nervous system caused by conventional viruses, perivascular leukocyte infiltrations do not play any part in the progression of spongiform encephalopathies and it is not known whether leukocyte infiltrations can be induced by unconventional slow viruses. Here we report that scrapie infection can indeed produce such lesions in mice.

Perivascular cuffs have occasionally been found in brains of patients with Kuru^{1,2}, but there are conflicting data on scrapie. Perivascular cuffs have been reported to be absent from⁸ histological sections of brains from scrapie-diseased animals, but present in intracerebrally (i.c.) injected scrapie- and mock-infected goats⁹, and in brains of sheep injected i.c. with scrapie agent¹⁰. However, no controls were used in the latter study. Zlotnik and Rennie¹¹ observed similar lesions in naturally infected sheep but also in control animals which were apparently healthy. Therefore, they did not consider these lesions to be associated with the disease.

We first studied scrapie in inbred STU mice¹² using mouse brain containing Chandler agent (47th passage; from Dr R. Kimberlin). When injected i.c.¹³ and passaged in our mice, this material gives incubation periods of ~150 (10^{-2} dilution of brain) to 240 (10^{-7} dilution) days. Examination of haematoxylin-eosin-stained sections of brains from mice infected i.c. with scrapie material revealed perivascular cuffs in some sections of a

few animals at the clinical stage of the disease (Fig. 1a). We found these lesions only when the whole brain was sliced and five 10- μ m sections taken every 200 μ m. We therefore used this technique to analyse brains of age-matched, mock-infected and scrapie-infected animals for perivascular cuffs. Aliquots (50 μ l) of a 10^{-2} – 10^{-3} dilution of brains from scrapie-diseased or healthy mice were injected intraperitoneally (i.p.), to avoid any possible immune reaction to brain injury. The mice developed symptoms of clinical scrapie 170–200 days after infection.

In the first experiment, only animals infected with scrapie material were used and histology showed that 5/10 developed perivascular cuffs (Table 1). Four animals were scored as positive in the advanced incubation period of the disease, at days 150–192. Surprisingly, however, one animal was scored as positive at day 73 after infection when less than one-half of the incubation period had passed and clinical and histological diagnosis of scrapie was not possible. However, titres of the scrapie agent have been reported at this stage for Compton white mice¹⁴.

In experiment 2, coded brains of control and infected mice were examined 'blind'—without the histologist knowing the number of infected brains; 4 of 14 brains of scrapie-infected animals were scored as positive whereas no control animals showed perivascular cuffs. Again, in one mouse brain infiltrates were detected in the early phase of the incubation period, at day 63.

Experiment 3 examined only advanced cases of scrapie, with the result that even in severely sick animals, only a limited number of mice (4/10) showed lesions. However, note that our technique of serial sectioning allowed histological examination of only one-fifth of the total brain. Control animals were free of infiltrations.

We have no evidence that the perivascular cuffs observed could be related to a possible contagious agent in the mouse stock as: (1) STU mice are no more susceptible to infections than conventional inbred mouse strains; (2) infected and mock-infected animals were kept in the same room; and (3) the experiments were performed at different times as indicated in Table 1. (4) No mice died during the experiment and none looked sick before clinical symptoms of scrapie appeared. (5) In experiments 2 and 3, the spleens were removed and the weights of single spleens recorded. There was no indication of a splenomegaly related to the perivascular cuffs.

Because of the long incubation period, even after i.c. injection, and as perivascular cuffs also developed in an animal

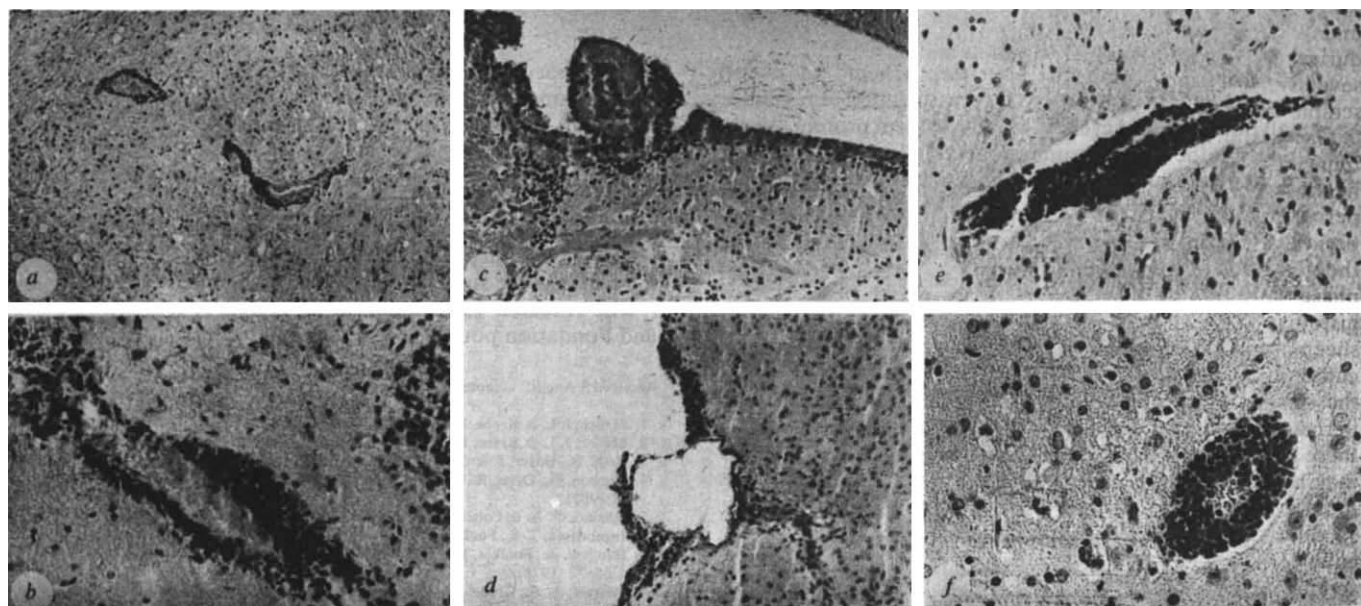


Fig. 1 Examples of meningeal leukocyte infiltrations and perivascular cuffs in scrapie-infected mouse brains. *a*, Perivascular cuffs 138 days after i.c. injection ($\times 58$); and *b*, 63 days after i.p. injection ($\times 133$). *c*, Meningeal infiltration 73 days after i.p. injection ($\times 85$); and *d*, 147 days after i.p. injection ($\times 85$). *e*, Perivascular cuffs 163 days after i.p. injection ($\times 133$); and *f*, 204 days after i.p. injection ($\times 217$).

Table 1 Leukocyte infiltrations in mouse brain after i.p. injection of 10^{-2} – 10^{-3} dilutions of brain material from healthy or scrapie-diseased mice

Expt	No. of animals injected		Days post-injection on which animals were killed for histology	No. of animals with leukocyte infiltrations	
	Control	Scrapie		Control	Scrapie
1 (1979)	—	10	25, 56, 73(+), 120, 132, 150(+), 170(2)(+), 178(+), 192(+)	—	5/10
2 (1980)	21	14	20, 35, 42, 56, 63(+), 70, 77, 91, 105, 119, 133, 147(+), 161(+), 176(+)	0/21	4/14
3 (1980)	4	14	190, 204	0/4	4/14
4 (1981)	12*	12*	65(12), 71(+)(12)	0/12	2/12

Numbers in parentheses in column 4 indicate numbers of animals taken for histology. Animals in which perivascular cuffs were found are indicated by (+). All histology was done using coded material.

* Animals received brain material treated with formalin.

which received an i.c. inoculum of a 10^{-7} dilution of scrapie brain material, it is unlikely that our passaged material could have been contaminated by another contagious agent. Nevertheless, we tried to exclude this possibility by exploiting the high degree of resistance of the scrapie agent to formaldehyde^{15,16}. To destroy any other infectious agents, a low speed (500g for 5 min) supernatant of a 10^{-1} brain homogenate was pretreated with 8% formalin at room temperature for 16 h and then injected (10^{-3} dilution) into mice. As in experiment 2, coded brains were scored for perivascular cuffs 65 and 71 days after infection, when diagnosis of scrapie was not possible. All controls were free of perivascular cuffs, whereas 2/12 animals of the group which had received scrapie-infected material were positive. Examples of leukocyte infiltrations at various times after i.p. injection are shown in Fig. 1; we have selected mild (Fig. 1c, d) and more severe (b, e, f) infiltrations to indicate the range of our scoring. Perivascular cuffs have been observed mainly in the hippocampus and thalamus at very low frequency during the routine examination of many thousands of brains of several mouse strains infected with various different strains of scrapie (H. Fraser, personal communication).

Our results thus show that leukocyte infiltrations can be induced by an unconventional slow virus and that brains of healthy animals are free from these lesions; that in some cases perivascular cuffs can be detected relatively early in the latent period, when diagnosis of the disease is not possible, but that they do not occur in all infected animals, even in the clinical phase when the disease can be diagnosed by symptoms and histology. Thus it is possible that the apparently healthy control sheep observed by Zlotnik and Rennie¹¹ to contain perivascular cuffs but showing no clinical symptoms were animals harbouring the scrapie agent. Although our technique could not detect perivascular cuffs in all scrapie-infected animals, we suggest that these lesions are not incidentally associated with the disease but may be a consequence of a very weak virus-induced autoimmune response, as has been reported for Kuru and Creutzfeldt-Jacob diseases¹⁷. However, it is also possible that the perivascular cuffs are caused by an immune reaction to a product of the scrapie agent which has extremely low antigenicity. Furthermore, our results indicate that this may also be the case for diseases of the central nervous system in man, in which perivascular cuffs are seen occasionally, but which are not known to be caused by an infectious agent.

We thank Miss J. Haase for technical assistance.

Received 6 July; accepted 5 October 1981.

- Gibbs, C. J. Jr & Gajdusek, D. C. *Perspect. Virol.* **10**, 161–198 (1978).
- Lampert, P. W., Gajdusek, D. C. & Gibbs, C. J. Jr *Am. J. Path.* **68**, 626–651 (1972).
- Kimberlin, R. H. *Sci. Prog. Oxf.* **63**, 461–481 (1976).
- Fraser, H. in *Progress in Neurological Research* (eds Behan, P. O. & Rose, F. C.) 194–210 (Pitman Medical, London, 1979).
- Chandler, R. L. *Lancet* **i**, 1378–1379 (1961).
- Hadlow, W. J. *Lancet* **ii**, 289–290 (1959).
- Lampert, P., Hooks, J., Gibbs, C. J. Jr & Gajdusek, D. C. *Acta Neuropath.* **19**, 81–93 (1971).
- Eklund, C. M., Hadlow, W. J. & Kennedy, R. C. *Proc. Soc. exp. Biol. Med.* **112**, 974–979 (1963).
- Hadlow, W. J. *Res. Vet. Sci.* **2**, 289–314 (1961).
- Wilson, D. R., Anderson, R. D. & Smith, W. J. *J. comp. Path.* **60**, 267–282 (1950).
- Zlotnik, I. & Rennie, J. C. *J. comp. Path.* **67**, 30–36 (1957).
- Schäfer, W. *Z. Naturf.* **34 c**, 306–309 (1979).
- Kimberlin, R. H. *J. Neurochem.* **19**, 2767–2778 (1971).
- Kimberlin, R. H. & Walker, C. A. *J. comp. Path.* **89**, 551–562 (1979).
- Pattison, J. H. *J. comp. Path.* **75**, 159–164 (1965).
- Hunter, G. D., Gibbons, R. A. & Kimberlin, R. H. *J. comp. Path.* **79**, 101–108 (1969).
- Sotelo, J., Gibbs, C. J. Jr & Gajdusek, D. C. *Science* **210**, 190–193 (1980).

Immunization against blood-stage rodent malaria using purified parasite antigens

Anthony A. Holder & Robert R. Freeman

Department of Immunochemistry, The Wellcome Research Laboratories, Langley Court, Beckenham, Kent BR3 3BS, UK

We have reported the production of several hybridoma lines secreting monoclonal antibodies recognizing antigens of the blood forms of a rodent malaria parasite, *Plasmodium yoelii yoelii*^{1,2}. Line WIC 25.77 secretes an IgG2a (antibody 25.77) which reacts specifically with merozoites in the indirect immunofluorescence (IIF) assay, and which was protective on passive transfer. Another of the hybridoma lines, WIC 25.1, secretes an IgG2a (antibody 25.1) which reacts with an antigen apparently associated with the membranes of schizonts, but this antibody was not protective on passive transfer. We have used our monoclonal antibodies to identify and purify antigens of *P. y. yoelii*, and we report here that mice immunized with either a 235,000 apparent molecular weight (MW) merozoite-specific protein or a 230,000 MW schizont protein and its proteolytic derivatives were protected against infection with *P. y. yoelii*.

Schizonts were obtained by centrifugation of *P. y. yoelii*-infected mouse erythrocytes through a layer of Histopaque, and were metabolically labelled *in vitro* with ³⁵S-methionine. The labelled schizonts (including a subpopulation of segmenters containing mature merozoites) were solubilized in a buffer containing a nonionic detergent. Immunoprecipitation of the extract followed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) of the washed immune complexes and fluorography revealed that antibody 25.77 reacted with a single polypeptide of 235,000 MW whereas antibody 25.1 precipitated eight major polypeptides and several minor bands (Fig. 1a, b). The apparent MWs of the major polypeptides were estimated to be 230,000, 197,000, 160,000, 151,000, 126,000, 90,000, 56,000, and 28,000.

Preparative scale purification of these antigens was achieved by monoclonal antibody affinity chromatography. Washed *P. y. yoelii*-parasitized mouse erythrocytes were extracted in a buffer containing 1% (v/v) NP40 and protease inhibitors. After high speed centrifugation the supernatant was applied sequentially to two columns containing 25.77 antibody-Sepharose and 25.1 antibody-Sepharose, respectively. Specific elution of bound protein from the 25.77 antibody column was achieved at pH 8.0 with 2 M potassium thiocyanate and 50 mM diethylamine, pH 11.5, was used for specific elution of protein bound to the 25.1 antibody column. After a second cycle of absorption and elution the purity of the protein preparations was assessed by SDS-PAGE. Essentially, the only polypeptides present were

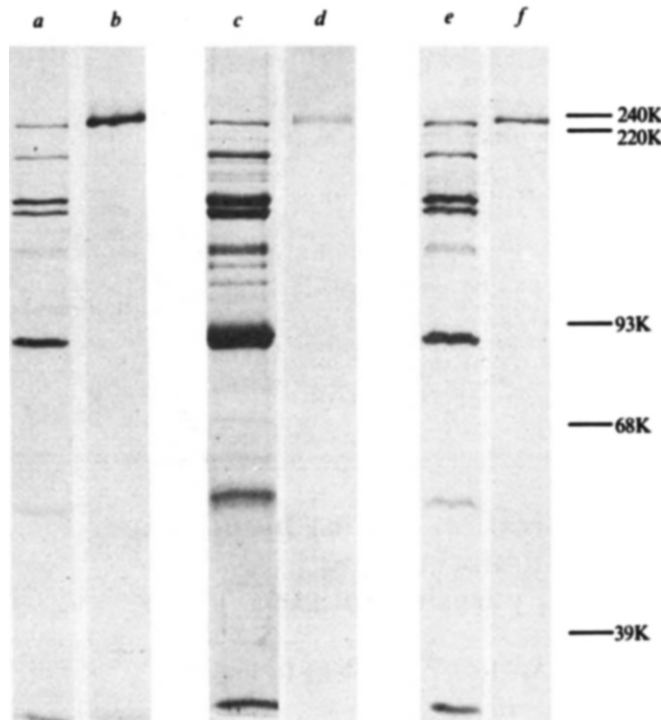


Fig. 1 *P. y. yoelii* polypeptides analysed by SDS-PAGE⁷ in 7.5% gels. In tracks *a*, *b* and *e*, *f*, the polypeptides are labelled with ³⁵S-methionine and visualized by fluorography, and in tracks *c*, *d*, they are stained with Coomassie blue. Track *a*, products immunoprecipitated with antibody 25.1; track *b*, immunoprecipitation with antibody 25.77; track *c*, protein purified by monoclonal antibody affinity chromatography using antibody 25.1; track *d*, protein purified by monoclonal antibody affinity chromatography using antibody 25.77; track *e*, ³⁵S-labelled polypeptides immunoprecipitated by serum from mice immunized with the protein preparation shown in track *c*; track *f*, products immunoprecipitated by serum from mice immunized with the protein preparation shown in track *d*. A schizont-enriched fraction of parasitized erythrocytes, prepared by centrifugation through a layer of Histopaque (Sigma)⁸, was labelled with ³⁵S-methionine as previously described². Labelled (for immunoprecipitation) or unlabelled (for affinity chromatography) parasitized erythrocytes were extracted on ice in at least 5 volumes of a buffer containing 50 mM Tris-HCl pH 8.0, 5 mM EDTA, 5 mM EGTA, 5 mM iodoacetamide, 1 mM phenylmethyl sulphonyl fluoride, 0.1 mM TLCK, and 1% (v/v) NP40. Insoluble material was removed by centrifuging at 100,000g for 45 min at 4°C. Formalin-fixed Cowan I strain *Staphylococcus aureus* organisms were used to precipitate immune complexes⁹. For fluorographic visualization, gels were treated with En³Hance (NEN) before exposure at -70°C using Kodak X-Omat H film. Monoclonal antibodies were purified from ascitic fluid from BALB/c mice carrying hybridoma lines WIC 25.1 and WIC 25.77 using Protein A-Sepharose¹⁰. For preparative scale protein purification (tracks *c*, *d*), an extract from 5×10^{11} *P. y. yoelii*-parasitized erythrocytes was applied sequentially to 10-ml columns of CNBr-activated Sepharose 4B (Pharmacia), coupled with antibody 25.1 or antibody 25.77 according to the manufacturer's recommended method, and equilibrated with 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM EGTA, 1% (v/v) NP40, at 4°C. The 25.77 antibody-Sepharose column was washed with equilibration buffer containing 0.5 M NaCl then with 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM EGTA, 0.01% NP40, and then bound protein was eluted using 2 M KSCN in the same buffer. The eluate was dialysed against 0.1% NP40 and reappplied to the column. Retained material was re-eluted using 2 M KSCN,

0.01% NP40, concentrated and dialysed. The 25.1 antibody-Sepharose column was washed with equilibration buffer containing 0.5 M NaCl, then with buffer containing 0.1% NP40. Bound material was eluted with 50 mM diethylamine pH 11.5, 0.1% NP40. The eluate was dialysed against the pH 8.0 buffer containing 0.1% NP40, and applied a second time to the re-equilibrated column. The column was washed with 10 mM Tris-HCl pH 8.2, 1 mM EDTA, 1 mM EGTA, 0.5% (w/v) sodium deoxycholate, and the retained material was eluted with 50 mM diethylamine pH 11.5, 0.5% sodium deoxycholate, concentrated and dialysed. 400 µg of protein was routinely recovered from the 25.77 antibody-Sepharose column, and the 25.1 antibody-Sepharose column yielded some 1,600 µg of protein. Antisera against the purified *P. y. yoelii* proteins for immunoprecipitation analysis (tracks *e*, *f*) were raised in BALB/c mice as described in the Fig. 4 legend. Molecular weight markers were human spectrin heterodimer (240,000 and 220,000), phosphorylase *b* (93,000), bovine serum albumin (68,000) and aldolase (39,000).

identical to those previously detected in the monoclonal antibody immunoprecipitates (Fig. 1*c*, *d*).

The relationship between the polypeptides recognized by antibody 25.1 was revealed by pulse-chase labelling of *P. y. yoelii* schizonts with ³⁵S-methionine *in vitro*, followed by solubilization and immunoprecipitation of samples at various time intervals after labelling. The label was initially

incorporated into the 230,000 MW polypeptide, then during further incubation it sequentially appeared in the smaller polypeptides, with a concomitant reduction in the amount of label remaining in the 230,000 MW component (Fig. 2). Thus, the native antigen recognized by antibody 25.1 is a 230,000 MW protein and the smaller polypeptides are proteolytic fragments of it, each retaining the antigenic determinant recognized by the

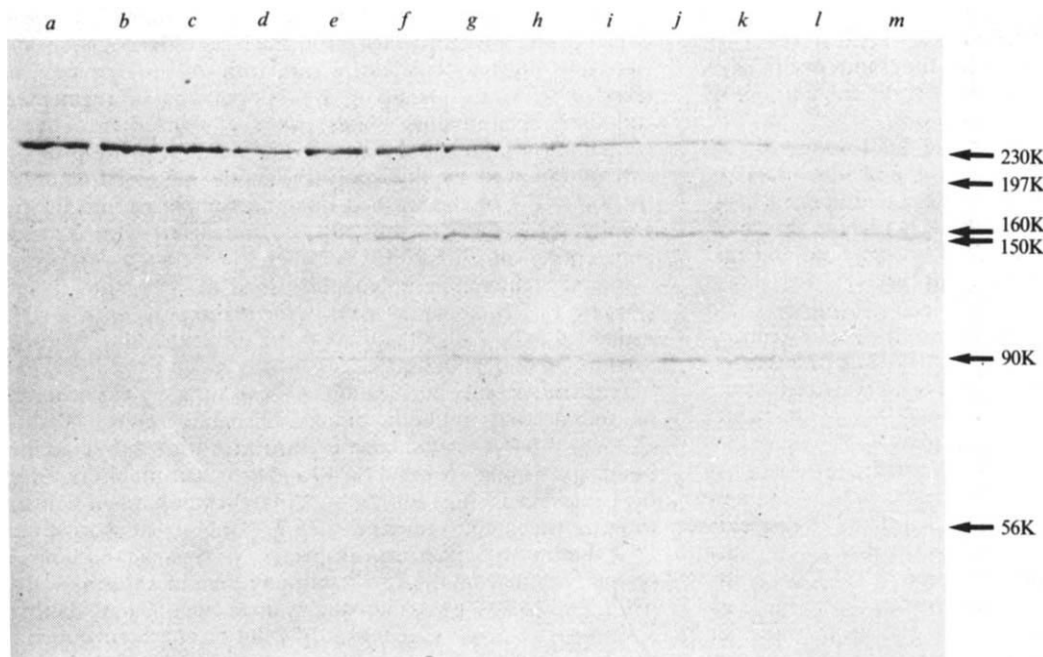
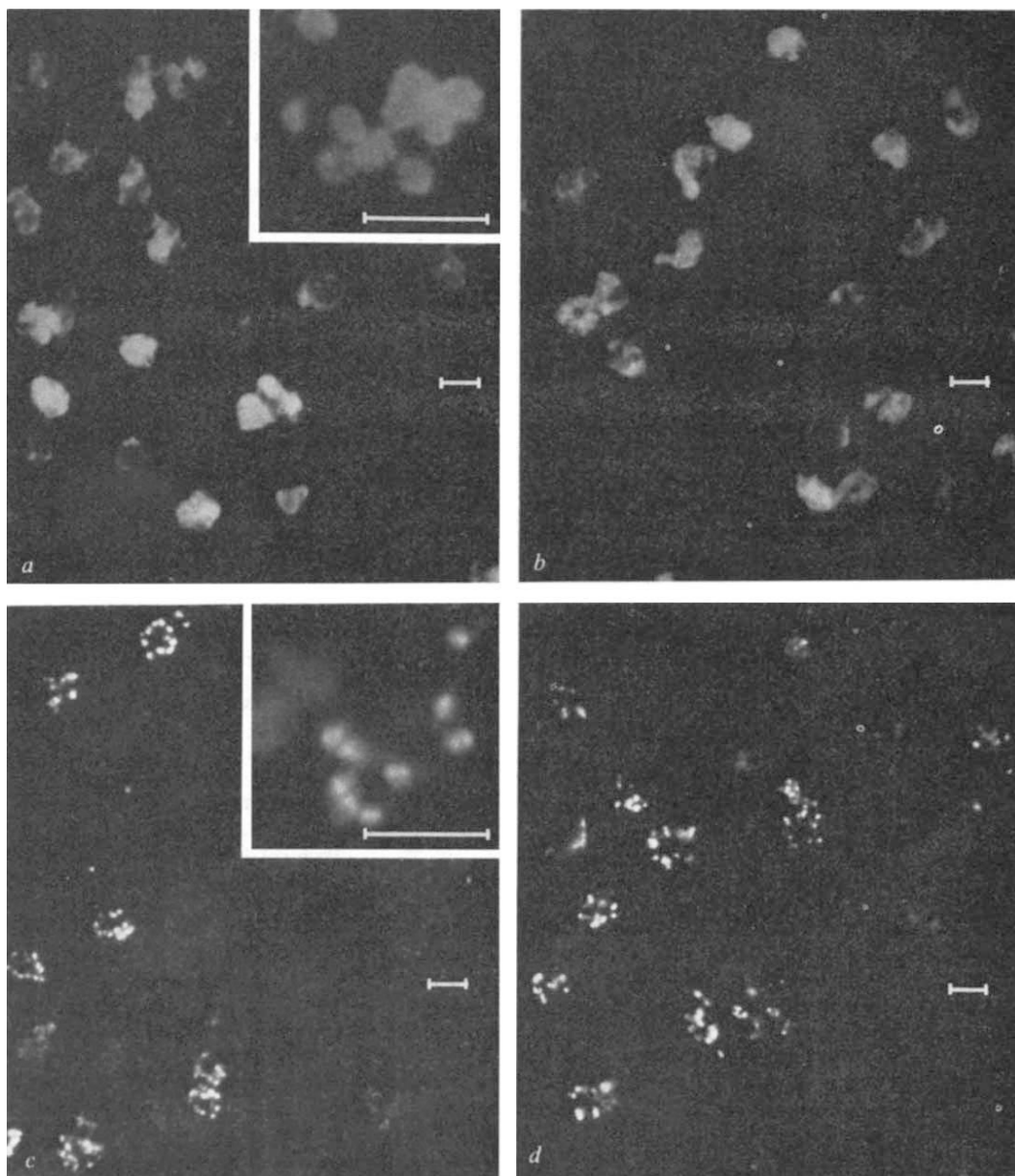


Fig. 2 Processing of the 230,000 MW schizont protein during incubation of *P. y. yoelii*-parasitized erythrocytes *in vitro*. Parasitized erythrocytes (4×10^8) were incubated at 37°C in 8.0 ml of methionine-free RPMI 1640 (Gibco) tissue culture medium supplemented with $10 \mu\text{Ci ml}^{-1}$ ³⁵S-methionine (Amersham). After 10 min, 400 µl of 50 mM methionine was added, and 0.5-ml aliquots of cell suspension were removed at the times indicated. The cells were washed with ice-cold RPMI 1640 and then frozen at -80°C. The frozen cell pellets were thawed by addition of 0.5 ml of 50 mM Tris-HCl pH 8.0, 5 mM EDTA, 5 mM EGTA, 5 mM iodoacetamide, 1 mM phenylmethyl sulphonyl fluoride, 0.1 mM TLCK, 1% (v/v) NP40, and used for immunoprecipitation using antibody 25.1. After SDS-PAGE on

a 7.5% gel, ³⁵S-labelled polypeptides were detected by fluorography at -70°C. The estimated MWs of the immunoprecipitated polypeptides are indicated (see Fig. 1). Duration of chase (min): *a*, 0; *b*, 10; *c*, 20; *d*, 40; *e*, 60; *f*, 90; *g*, 120; *h*, 150; *i*, 180; *j*, 240; *k*, 300; *l*, 360; *m*, 420.

Fig. 3 IIF staining reactions on acetone-fixed *P. y. yoelii* YM antigen smears, using serum from BALB/c mice carrying the hybridoma lines WIC 25.1 (a) and 25.77 (c), or using serum from mice immunized with the 230,000 MW *P. y. yoelii* protein and its fragments (b) and the 235,000 MW *P. y. yoelii* protein (d), as described in Fig. 4 legend. The IIF assay was performed as previously described¹, using immunoblot-purified fluorescein isothiocyanate-labelled rabbit anti-mouse IgG. Inset a, IIF reaction of antibody 25.1 on a mature segmented schizont. Note localization of staining to a membrane surrounding individual merozoites. Inset c, IIF reaction of antibody 25.77 on a mature segmented schizont. Note localization of staining to a small area on each merozoite. Scale bar, 5 μ m.



monoclonal antibody. Amino acid sequence homology between the major immunoprecipitated polypeptides was confirmed by two-dimensional chymotryptic peptide mapping (data not shown). *In vitro*, degradation or processing of the native protein began soon after its synthesis and was extensive after 7 h (Fig. 2). The results indicate that the processing occurred during incubation of the intact schizonts rather than after cell lysis and solubilization. As a consequence of the phenomenon, our purified 25.1 protein preparations always contained fragments of the 230,000 MW protein as well as the intact protein.

In the IIF test, antibody 25.1 reacted with free merozoites in suspension, indicating that the 230,000 MW protein (or fragments of it) is located on the merozoite surface. Using acetone-fixed schizonts, the IIF staining by antibody 25.1 suggests that the 230,000 MW protein is synthesized before merozoite formation and is associated specifically with the plasma membrane of the developing intracellular parasite (Fig. 3a), rather than with the host erythrocyte plasma membrane as suggested previously². In early schizonts the antigen seems to be localized at a spherical membrane enclosing the parasite, but following segmentation and formation of the residual body, the staining is evenly distributed over each merozoite (Fig. 3a, insert).

Antibody 25.77 did not react with free merozoites in suspension in the IIF test, and with acetone-fixed schizonts the reaction

was localized to a region within each merozoite (Fig. 3c, insert). It seems, therefore, that the 235,000 MW protein is not on the surface, but may be associated with intracellular organelles of the mature merozoite.

Groups of BALB/c mice were immunized with the two *P. y. yoelii* protein preparations on three occasions. One week after the third immunization, the mean serum titres were 1 in 10,240 (merozoite-specific) for the group immunized with the 235,000 MW protein, 1 in 10,240 (schizont-specific) for the group immunized with the 230,000 MW protein and its fragments, and 1 in 40 (nonspecific) for the control group.

The sera from the immunized mice immunoprecipitated the same polypeptides as did the monoclonal antibodies (Fig. 1e, f), and produced IIF staining patterns indistinguishable from those produced by the monoclonal antibodies (Fig. 3), confirming the immunogenicity and purity of the parasite protein preparations.

The groups of mice immunized as described were challenged by intravenous (i.v.) injection of 10,000 *P. y. yoelii* YM-parasitised erythrocytes. In the control group, parasitaemia increased rapidly and all mice in this group died on day 8 after challenge (Fig. 4). In contrast, the mice immunized with the 230,000 MW protein and its fragments showed relatively low-grade parasitaemias which were cleared from blood smears by day 10, and all mice in this group survived. In the group immunized with the 235,000 MW protein, parasitaemia was initially suppressed

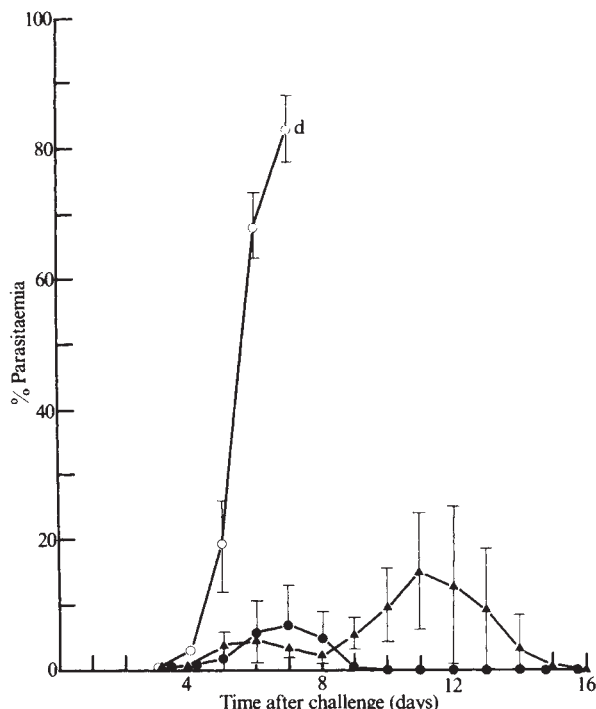


Fig. 4 Mean parasitaemias in groups of five BALB/c mice challenged with 10,000 *P. y. yoelii* YM-parasitized erythrocytes, after immunization with the 230,000 MW *P. y. yoelii* protein and its fragments (●), or with the 235,000 MW *P. y. yoelii* protein (▲) or saline (○). The *P. y. yoelii* proteins were purified as described in Fig. 1 legend, and the immunization protocol was as follows: 12 µg of protein in Freund's complete adjuvant (FCA) given intraperitoneally (i.p.) on days 0 and 35, and 20 µg of protein in saline given i.v. on day 50. In the control group, 0.1 ml of saline in FCA was given i.p. on days 0 and 35, and 0.1 ml of saline was given i.v. on day 50. The challenge infection was given i.v. on day 60. Parasitaemia was scored daily by microscopic counting of parasitized erythrocytes in smears of tail blood stained with Giemsa's stain. Mean survival time in the control group is indicated (d). The vertical bars indicate standard deviations.

relative to controls, but after day 8 an influx of reticulocytes into the circulation was coincident with an increase in parasitaemia, with all parasites confined to reticulocytes. However, parasitaemias in this group were cleared from blood smears by day 16, and all the mice immunized with the 235,000 MW protein survived.

Immunization with the 230,000 MW schizont antigen and its fragments induced an immune response equally effective against parasites in reticulocytes and in mature erythrocytes. In immunized mice, challenge infection developed normally for the first 3 days, was then suppressed relative to controls, and cleared over the following 7 days. This observation indicates that resistance to infection in these mice was not due simply to the serum antibody response to immunization. The kinetics of challenge infection were similar to those in mice immunized with whole formalin-fixed, free *P. y. yoelii* parasites described by Playfair *et al.*³. These workers have implicated both humoral and cell-mediated immune responses in the mechanism of parasite clearance in mice immunized with whole parasites^{4,5} and it will be important to analyse further the mechanisms mediating immunity in mice immunized with the purified *P. y. yoelii* proteins.

The ability of *P. y. yoelii* strains to invade mature erythrocytes is an important factor in their virulence⁶. As had previously been observed² after passive transfer of antibodies 25.77 and 25.37, challenge infections in mice immunized with the 235,000 MW merozoite antigen were confined to reticulocytes, and the infections were non-lethal. In unimmunized controls, on the other hand, mature erythrocytes also became extensively parasitized, leading to the deaths of these mice. How the 235,000 MW antigen might be involved in host cell specificity and virulence of *P. y. yoelii* is not understood.

Other protective *P. y. yoelii* antigens may exist, and it should be possible, using the approach described here, to identify and

characterize them. To extend the present findings, it will be important to determine whether proteins analogous to the 230,000 MW and 235,000 MW antigens of *P. y. yoelii* are synthesized by other malaria parasites, in particular the human malaria parasite, *Plasmodium falciparum*.

We thank Lynn Bushby, Lynne Davey and Sheila Graves for technical assistance and Dr G. A. M. Cross for helpful discussions.

Received 3 July; accepted 13 October 1981.

- Freeman, R. R., Trejdosiewicz, A. J. & Cross, G. A. M. *Nature* **284**, 366–368 (1980).
- Freeman, R. R., Holder, A. A., Trejdosiewicz, A. J. & Cross, G. A. M. in *The Host-Invader Interplay* (ed. Van den Bossche, H.) 121–124 (Elsevier, Amsterdam, 1980).
- Playfair, J. H. L., De Souza, J. B. & Cottrell, B. J. *Immunology* **33**, 507–515 (1977).
- Playfair, J. H. L. & De Souza, J. B. *Parasite Immun.* **1**, 197–208 (1979).
- Playfair, J. H. L., De Souza, J. B., Dockrell, H. M., Agomo, P. U. & Taverne, J. *Nature* **282**, 731–734 (1979).
- Walliker, D., Sanderson, A., Yoeli, M. & Hargreaves, B. J. *Parasitology* **72**, 183–194 (1976).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).
- Davidson, W. F. & Parish, C. R. *J. Immun. Meth.* **7**, 291–300 (1975).
- Kessler, S. W. *J. Immun.* **115**, 1617–1624 (1975).
- Ey, P. L., Prowse, S. J. & Jenkin, C. R. *Immunochimistry* **15**, 429–436 (1978).

Complete *in vitro* maturation of *Plasmodium falciparum* gametocytes

T. Ifediba & J. P. Vanderberg

Division of Parasitology, Department of Microbiology, New York University Medical School, New York, New York 10016, USA

The form of malaria caused by *Plasmodium falciparum* is probably the most important infectious disease of man. In tropical Africa alone, where malaria affects almost the entire population, it has been estimated that every year the disease causes the death of 1 million children under 14 yr old¹. The mosquito vector of malaria is infected when it ingests mature gametocytes in blood taken from a human carrier of the disease. Development of the parasite continues in the mosquito and culminates with the sporozoite stage, which initiates a new infection when injected by the mosquito into a further human host. Research on *P. falciparum* sporozoites has been severely limited by a general lack of availability of suitable patients with gametocytes. It was hoped that the recently developed continuous cultivation of *P. falciparum*² would solve this, as gametocytes are often produced during cultivation. However, we and others soon found that these gametocytes failed to mature in culture^{3,4}, or at best did so only rarely and unpredictably^{5–7}. We report here that addition of hypoxanthine to the culture medium permits the production of mature, infectious *P. falciparum* gametocytes on a regular basis.

Gametocyte production involves two distinct processes: an inductive process that commits certain asexual parasites to become gametocytes, and a developmental process that supports the full differentiation of these committed cells³. Gametocyte induction has been shown to occur more readily when the cultures are maintained for extended periods of time without the addition of fresh uninfected erythrocytes^{4,6}. We have found that this induction process is enhanced when the erythrocyte concentration is reduced for part of the culture period, and when old cultures with mature gametocytes are used to initiate new cultures (Table 1). When the erythrocyte concentration was reduced to 6%, the resulting gametocytaemia was almost double that observed for an erythrocyte concentration of 12%. Interestingly, the absolute numbers of gametocytes per culture flask remained about the same in both cases.

Gametocytes grown in these conditions, and with added hypoxanthine (50 µg ml⁻¹), developed to functional maturity (as determined by their infectivity for mosquitoes (Tables 2–4)). When the gametocytes were collected 12–14 days after initiation of the cultures, mosquitoes in each of 18 groups that fed on

Table 1 Effect of erythrocyte concentration on production *in vitro* of *Plasmodium falciparum* gametocytes

Strain	Erythrocyte concentration	No. of cultures	Gametocytaemia*
Honduras	12%	72	1.5 ± 0.5%
	6%	74	2.7 ± 0.5%
Z	12%	23	1.4 ± 0.6%
	6%	24	2.9 ± 0.6%
FCN-2	12%	28	1.0 ± 0.5%
	6%	26	2.3 ± 0.6%

All cultures were initiated at 0.2% parasitaemia and 12% haematocrit in RPMI-1640 with HEPES², 15% (v/v) human serum and 50 µg ml⁻¹ hypoxanthine (Sigma). Human erythrocytes were used within a few days after collection. Cultures were grown at 37 °C in tissue culture flasks (75 cm²) in a total volume of 20 ml per flask, in an atmosphere of 3% O₂, 2% CO₂ and 95% N (ref. 15). Medium was changed daily by removal of spent medium and addition of fresh medium prewarmed to 37 °C. No new erythrocytes were added during the culture period. At 7 days the haematocrit of some cultures was reduced to 6%, and daily feeding of all cultures continued with the same medium, but with the human serum component reduced to 10%. Gametocytaemia was determined at 12–14 days, and these mature cultures were used to initiate new gametocyte cultures. The Honduras strain was from CDC, Z was obtained from NIH and FCN-2 was isolated from blood of a patient from Bangkok.

* Number of gametocytes (±s.d.) per 100 erythrocytes, determined from Giemsa-stained smears.

them developed midgut infections consisting of substantial numbers of oocysts. In each of these groups, the percentage of mosquitoes that went on to develop salivary gland infections of sporozoites was comparable to the percentage with oocyst infections. Sporozoites collected from the salivary glands appeared normal, and had motility typical of *P. falciparum* sporozoites⁸.

Complete gametocyte maturation did not normally occur in parallel control cultures without added hypoxanthine (Table 2). Gametocytes in these control cultures appeared morphologically normal, but were generally not infective for mosquitoes. In a single case, a mosquito fed on gametocytes from these control cultures was found to have one oocyst in its midgut.

Asexual, erythrocytic malarial parasites are known to require exogenous sources of purines for their metabolism, but rely on *de novo* synthesis of pyrimidines⁹. Studies of the sporogenic

Table 2 Effects of hypoxanthine on gametocyte maturation in infected vectors

Expt	No. of mosquitoes infected with oocysts	
	Hypoxanthine (50 µg ml ⁻¹)	No hypoxanthine
1	11/17 (65%)	0/15 (0%)
2	8/20 (40%)	0/20 (0%)
3	14/18 (78%)	1/17* (6%)
4	13/20 (65%)	0/13 (0%)
5	10/16 (63%)	0/15 (0%)

Gametocytes (Honduras strain) were cultured as described in Table 1 legend, with or without hypoxanthine added to the culture medium, and collected 13–14 days after initiation of the cultures. Cells in the culture flask were sedimented, after which 2–3 vol of uninfected human erythrocytes were added together with sufficient human serum to give a 50% haematocrit, and fed to *Anopheles stephensi* mosquitoes using a membrane feeder¹⁶. Mosquitoes were maintained according to standard procedures¹⁷ and received a blood meal from an uninfected hamster on day 6. Midgut infections were determined 10–14 days after the infective feeding. Values shown are the results of individual experiments.

* Infection with single oocyst.

Table 3 Infectivity of gametocytes collected at different times after initiation of cultures

Age of gametocytes fed to mosquitoes (no. of days in culture)	No. of mosquitoes infected with oocysts					
	Honduras strain			FCN-2 strain		
11	6/14 (43%)					
12	8/14 (57%)	8/8 (100%)	5/12 (42%)	7/8 (88%)	10/13 (77%)	
13	4/12 (33%)			6/10 (60%)		
14	20/20 (100%)	25/27 (93%)	11/22 (50%)	10/24 (42%)	25/27 (83%)	4/12 (33%)
15	30/56 (54%)	0/15 (0%)	4/22 (18%)	7/14 (50%)	8/18 (44%)	2/11 (18%)
16			0/9 (0%)			
19	14/28 (50%)	0/30 (0%)		8/15 (53%)		

Gametocytes were cultured with added hypoxanthine and fed to mosquitoes as described in legends to Tables 1 and 2. Midgut infections were determined 10–14 days after the infective feeding. Values shown are the results of individual experiments.

stages of the parasite suggest that a similar situation may hold for ookinetes¹⁰ and sporozoites¹¹. There is evidence⁹ that the key exogenous purine incorporated by the parasite is hypoxanthine. We have previously pointed out³ that RPMI-1640 is a relatively simple tissue culture medium that supplies all nutritional requirements for asexual parasites but that gametocytes may require additional growth factors for full development. For successful cultivation of asexual parasites, the host erythrocyte is obliged to supply nutrients for only the 48-cycle of the parasite's development within it, whereas the developing gametocyte occupies its host erythrocyte for 2 weeks or longer. For complete gametocyte maturation to occur, the depleted hypoxanthine reserves of the host erythrocyte apparently must be supplemented by an exogenous source. Our finding that hypoxanthine may be a limiting factor in the culture of *P. falciparum* gametocytes should now permit large-scale production of mature gametocytes for further immunization studies using gametes¹² or sporozoites^{13,14} of this species.

We thank R. Nawrot and M. Silcott for technical assistance. This study was supported by the US Agency for International Development under contract AID/ta-C-1486.

Table 4 Oocyst production in *Anopheles stephensi* mosquitoes fed on *Plasmodium falciparum* gametocytes grown *in vitro*

Age of gametocytes fed to mosquitoes (no. of days in culture)	Mean no. of oocysts per infected mosquito					
	Honduras strain			FCN-2 strain		
11	2.8					
12	3.3	7.9	4.8	102.5	120.1	
13	2.0			3.0		
14	64.0	47.0	9.0	6.5	34.0	1.5
15	31.6		3.0	5.5	25.4	2.0
19	4.5			5.0		

Gametocytes were cultured with added hypoxanthine and fed to mosquitoes as described in Tables 1 and 2. Oocysts were counted 10–12 days after the infective feeding. Values shown are the results of individual experiments.

Received 2 September; accepted 8 October 1981.

1. Maugh, T. H. *Science* **196**, 413–416 (1977).
2. Trager, W. & Jensen, J. B. *Science* **193**, 673–675 (1976).
3. Vanderberg, J. P., Weiss, M. M. & Mack, S. R. *Bull. Wild Hlth Org.* **55**, 377–392 (1977).
4. Jensen, J. B. *J. Protozool.* **26**, 129–132 (1979).
5. Carter, R. & Beach, R. F. *Nature* **270**, 240–241 (1977).
6. Carter, R. & Miller, L. H. *Bull. Wild Hlth Org. (Suppl.)*, **57**, 37–52 (1979).
7. Campbell, C. C., Chin, W., Collins, W. E. & Moss, D. M. *Trans. R. Soc. trop. Med. Hyg.* **74**, 668–669 (1980).
8. Vanderberg, J. P. *J. Protozool.* **21**, 527–537 (1974).
9. Sherman, I. W. *Microbiol. Rev.* **43**, 453–495 (1979).
10. Weiss, M. M. & Vanderberg, J. P. *J. Protozool.* **23**, 547–551 (1976).
11. Jacobs, R. L., Miller, L. H. & Koontz, L. C. *J. Parasit.* **60**, 340–343 (1974).
12. Carter, R. & Gwadz, R. W. in *Malaria* Vol. 3 (ed. Kreier, J. P.) 263–297 (Academic, New York, 1980).
13. Cochrane, A. H., Nussenzeig, R. S. & Nardin, E. H. in *Malaria* Vol. 3 (ed. Kreier, J. P.) 163–202 (Academic, New York, 1980).
14. Clyde, D. F., Most, H., McCarthy, V. & Vanderberg, J. P. *Am. J. med. Sci.* **266**, 169–177 (1973).
15. Scheibel, L. W., Ashton, S. H. & Trager, W. *Expl Parasit.* **47**, 410–418 (1979).
16. Rutledge, L. C., Ward, R. A. & Gould, D. J. *Mosquito News* **24**, 407–419 (1964).
17. Vanderberg, J. P. & Gwadz, R. W. in *Malaria* Vol. 2 (ed. Kreier, J. P.) 153–234 (Academic, New York, 1980).

Opposing roles for D-1 and D-2 dopamine receptors in efflux of cyclic AMP from rat neostriatum

J. C. Stoof* & J. W. Keabian

Experimental Therapeutics Branch, National Institute of Neurological and Communicative Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20205, USA

Two types of dopamine receptor whose stimulation affects cellular cyclic AMP have now been identified. In tissues as different as insect ganglia and mammalian neostriatum, stimulation of the dopamine receptor called D-1 increases formation of cyclic AMP^{1–6}, whereas stimulation of the second type of dopamine receptor (D-2)^{7–11}, first identified in the rat pituitary gland, reduces cyclic AMP formation. Recently, a receptor with the pharmacological properties of the D-2 receptor has also been found^{12–14} in the rat neostriatum; we show here that stimulation of this receptor is followed by a reduction in cyclic AMP formation induced by V stimulation with D-1 agonists.

Previous investigations of the dopaminergic regulation of neostriatal cyclic AMP formation have determined either adenylate cyclase activity of tissue homogenates² or cyclic AMP content of tissue slices¹⁵. However, we have used as a biochemical sign of dopamine receptor stimulation the dopaminergic enhancement of cyclic AMP efflux from slices of neostriatum because cyclic AMP efflux is a particularly sensitive indicator of hormone-stimulated cyclic AMP formation in the adrenal cortex^{16,17}. Blocks of rat neostriatum were superfused with Earle's balanced salt solution, to which were added isobutyl methylxanthine¹⁸ (1 mM) and bovine serum albumin (BSA; 2.5 mg ml⁻¹) in a 16-chamber superfusion apparatus described elsewhere¹⁹. After 60 min of superfusion, several fractions were collected for each chamber, and dopaminergic drugs were added to the medium and additional fractions collected. At the end of the experiment, the cyclic AMP content of each fraction was determined by radioimmunoassay.

SKF 38393 is an agonist equipotent (on a molar basis) to dopamine on the D-1 dopamine receptor^{4,20,21}, but substantially less potent (on a molar basis) when tested on the pituitary D-2 dopamine receptor^{11,20}. SKF 38393 caused a time- and dose-dependent increase in the efflux of cyclic AMP from blocks of rat neostriatum (Fig. 1a, b), the maximal effect being a 3.5-fold enhancement of cyclic AMP efflux which occurred after 25 min. This maximal effect on cyclic AMP efflux was greater than the maximal effect of the agonist on either the cyclic AMP content of neostriatal blocks (1.6-fold enhancement; see Fig. 1 legend) or

Table 1 Fluphenazine inhibits SKF 38393-stimulated efflux of cyclic AMP

Drug	Cyclic AMP efflux (% of basal efflux)
None	99.8 ± 12.7
SKF 38393 (1 µM)	304.0 ± 18.5
SKF 38393 (1 µM) + fluphenazine (1 µM)*	194.3 ± 12.3†
SKF 38393 (1 µM) + fluphenazine (10 µM)*	121.8 ± 2.4†

Experimental procedures were as described for Fig. 1a. Basal efflux of cyclic AMP was determined for three successive 10-min fractions collected between 60 and 90 min. Exposure to drugs began at 90 min. Cyclic AMP efflux in the final fraction (115–125 min) was expressed as mean % of basal efflux (±s.e.m.) of observations made on four superfusion chambers (as described for Fig. 1b).

* Fluphenazine was added 30 min before SKF 38393 (see ref. 29).

† $P < 0.001$ compared with SKF 38393 alone (Student's *t*-test).

on the adenylate cyclase activity (1.7-fold enhancement²⁰). SKF 38393 half-maximally enhanced cyclic AMP efflux at a concentration of 2×10^{-7} M (Fig. 1c), which is about the same as the concentration (10^{-7} M) reported half-maximally to activate neostriatal adenylate cyclase activity²⁰. Fluphenazine markedly attenuated the SKF 38393-stimulated efflux of cyclic AMP (Table 1). However (–)sulpiride, an antagonist on the pituitary D-2 dopamine receptor^{11,22} which has no antagonist action on the D-1 receptor²³, did not alter the basal efflux of cyclic AMP and induced a statistically insignificant potentiation of the SKF 38393-stimulated efflux (Fig. 2a). These results demonstrate that the efflux of cyclic AMP from blocks of rat neostriatum provides a biochemical sign of stimulation of the D-1 receptor in this tissue.

LY-141865, an agonist on the pituitary D-2 receptor^{24,25} which has no effect on the D-1 receptor in fish retina²⁵ or rat neostriatum (E. Frey, unpublished results), did not stimulate the efflux of cyclic AMP either in the presence or absence of (–)sulpiride (Fig. 2b). Dopamine, an agonist on both the D-1 (refs 1–6) and D-2 (refs 11, 26, 27) dopamine receptors, increased the efflux of cyclic AMP from blocks of rat neostriatum (Fig. 2c). Interestingly, (–)sulpiride markedly potentiated the dopamine-stimulated efflux of cyclic AMP whereas (+)sulpiride was ineffective (Fig. 2c). Furthermore, apomorphine, which also stimulates both dopamine receptors^{1–4,11,26,27}, enhanced the efflux of cyclic AMP only in the presence of (–)sulpiride (Fig. 2d).

LY-141865 reduced the SKF 38393-stimulated efflux of cyclic AMP, the maximal effect being a reduction of $57 \pm 8\%$ (data from four separate experiments; Fig. 3). This inhibitory effect was reversed by (–)sulpiride (Fig. 3). LY-141865 did not significantly inhibit the (–)isoprenaline-stimulated efflux of cyclic AMP: cyclic AMP efflux induced by (–)isoprenaline (1 µM) was 514.3 ± 33.8 (% of basal efflux) compared with $452.0 \pm 28.3\%$ for (–)isoprenaline (1 µM) + LY-141865 (1 µM).

Based on the results obtained with dopamine, apomorphine and LY-141865, we propose the existence in the neostriatum of a second dopamine receptor, regulating the efflux (and by inference the formation) of cyclic AMP. Stimulation of this second receptor inhibits the formation of cyclic AMP occurring as a consequence of stimulating the D-1 dopamine receptor. According to this interpretation, dopamine interacts with both types of receptor in the neostriatum, simultaneously stimulating and inhibiting the formation and efflux of cyclic AMP. (–)Sulpiride potentiates the dopamine-stimulated efflux by blocking the second dopamine receptor, removing the inhibitory constraint on the formation of cyclic AMP and thereby stimulating the observed efflux of cyclic AMP. Similarly, apomorphine stimulates both dopamine receptors; (–)sulpiride antagonizes the interaction between apomorphine and the second dopamine receptor, thereby permitting the interaction between apomorphine and the D-1 receptor to be expressed as

* To whom correspondence should be addressed at: Department of Neurology, Medical Faculty, Free University, van der Boechorststraat 7, 1081BT Amsterdam, The Netherlands.

Fig. 1 SKF 38393 stimulates efflux of cyclic AMP from blocks of rat neostriatum. *a*, Time course of cyclic AMP efflux. Tissue was superfused without drug (●) or with SKF 38393 (Δ , 3×10^{-8} M; $\square$, 3×10^{-7} M; $\circ$, 3×10^{-6} M). Exposure to SKF 38393 (thickened abscissa) began at 80 min. Data represent mean $\pm$ s.e.m. ($n = 4$) of the cyclic AMP content of superfusion medium collected during the indicated time periods. At the end of the experiment, tissue content of cyclic AMP was 312 ± 14 fmol per mg wet weight in the absence of SKF 38393 and 352 ± 6 , 428 ± 22 and 506 ± 32 fmol per mg wet weight in the presence of 3×10^{-8} M, 3×10^{-7} M and 3×10^{-6} M SKF 38393 respectively. In all experiments, brains of eight male Sprague-Dawley rats (250–300 g) were removed and placed in a special holder. Two successive 2 mm-thick coronal brain slices were made between the anterior planes A 7,000 and A 11,000 μ m (ref. 30). Neostriatal tissue was dissected from these two slices and cut twice with a McIlwain tissue chopper (300 μ m); before the second pass, the chopping surface was rotated 90°. The resulting blocks of neostriatal tissue ($2 \times 0.3 \times 0.3$ mm) were transferred to Earle's balanced salt solution (EBSS) and stored until neostriatal tissue had been obtained from all rats. Aliquots (20 mg wet weight) of the chopped neostriatal tissue were placed in each of 16 chambers of a superfusion apparatus¹⁹ and superfused at a rate of 0.1 ml min^{-1} with EBSS (37 °C) containing BSA (2.5 mg ml^{-1}) and 3-isobutyl-1-methylxanthine¹⁸ (1 mM). After 60 min of superfusion, two or three 1.0-ml fractions (as indicated in the legends) were collected from each chamber; drugs were added to the superfusion medium and additional fractions collected from each chamber. Although no antioxidant was included during the superfusion, decomposition (as determined by HPLC with electrochemical detection) of drugs known to be sensitive to oxidation, such as dopamine, (–)isoprenaline or apomorphine, was <10% during the 30 min of superfusion. Each data point on the ordinate represents the midpoint of the collection period (for example, the data obtained from fractions collected between 80 and 90 min are located at a superfusion time of 85 min). At the end of the experiment, the tissue was superfused for 20 min with 0.1 M HCl to extract cyclic AMP from the tissue. The cyclic AMP content of duplicate 100- μ l aliquots of each fraction was estimated by radioimmunoassay (RIA) as described elsewhere³¹ using RIA kits (NEN); the limit of detection of this RIA was 2.5 fmol per 100 μ l. *b*, Data from *a* transformed mathematically. The ability of SKF 38393 to enhance cyclic AMP efflux was calculated from data in *a* (the same symbols are used). Every chamber served as its own control; for each chamber, the basal efflux of cyclic AMP (that is, the mean amount of cyclic AMP in the fractions collected before drug addition) was calculated, and the amount of cyclic AMP in each subsequent fraction expressed as a percentage of the basal efflux. *c*, Dose-response relationship of the SKF 38393-stimulated efflux of cyclic AMP determined from data obtained in nine separate experiments; the SKF 38393-induced stimulation of cyclic AMP efflux was calculated as in *b*. The data represent SKF 38393-stimulated cyclic AMP efflux in the last collected fraction (for example, the fraction collected between 110 and 120 min in *a*). Values represent mean $\pm$ s.e.m.; n , indicated in parentheses, is the number of superfusion chambers in which a given concentration of SKF 38393 was tested.

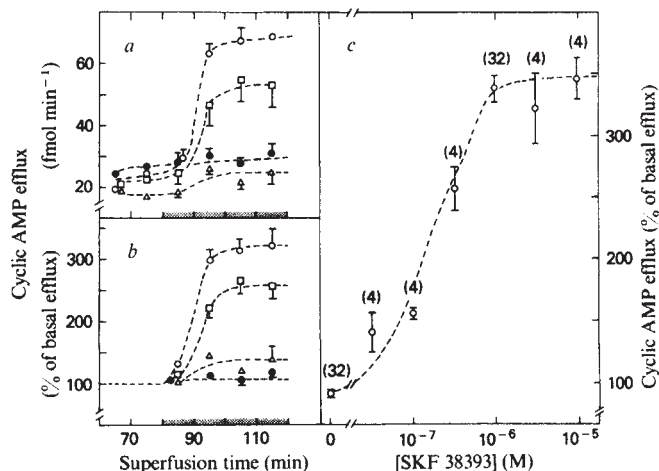
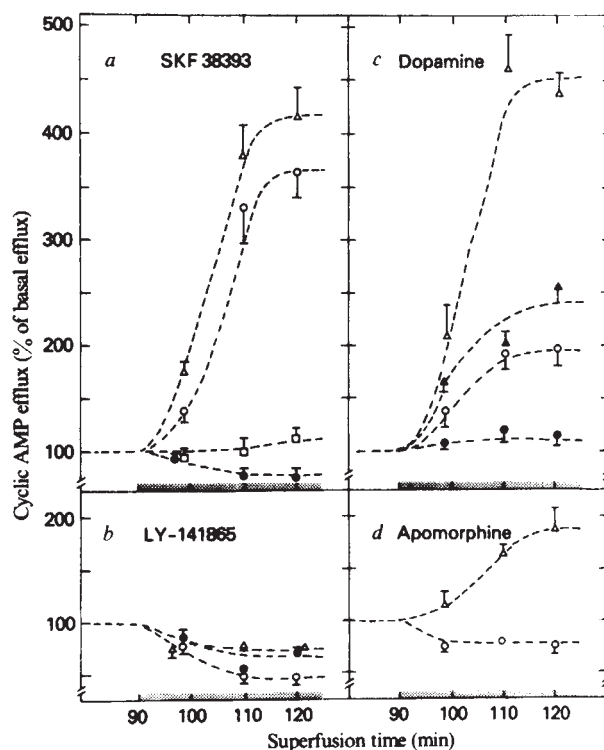


Fig. 2 Effects of dopaminergic agonists and sulpiride on efflux of cyclic AMP from blocks of rat neostriatum. Experimental procedures were as described for Fig. 1a. Basal efflux of cyclic AMP was determined from three successive 10-min fractions collected between 60 and 90 min. Exposure to drugs (thickened abscissa) began at 90 min. Cyclic AMP efflux was expressed as the mean % basal efflux ($\pm$ s.e.m.) of observations made on four superfusion chambers (as described for Fig. 1b). *a*, Tissue was superfused without drug (●), or with 10^{-6} M (–)sulpiride ($\square$), 10^{-6} M SKF 38393 ($\circ$) or 10^{-6} M (–)sulpiride and 10^{-6} M SKF 38393 (Δ). *b*, Tissue was superfused without drug (●) or with 10^{-6} M LY-141865 ($\circ$) or 10^{-6} M LY-141865 and 10^{-6} M (–)sulpiride (Δ). (LY-141865 is designated 'compound 38, R = pro' in ref. 24.) *c*, Tissue was superfused without drug (●) or with 8×10^{-6} M dopamine ($\circ$), 8×10^{-6} M dopamine and 2×10^{-6} M (–)sulpiride (Δ), or 8×10^{-6} M dopamine and 2×10^{-6} M (+)sulpiride ($\blacktriangle$). *d*, Tissue was superfused with 5×10^{-7} M apomorphine alone ($\circ$) or in combination with 10^{-6} M (–)sulpiride (Δ).



an enhanced efflux of cyclic AMP (although apomorphine can also antagonize the D-1 receptor^{3,21}, our experiments were not designed to demonstrate this property). The SKF 38393-stimulated efflux of cyclic AMP was not significantly potentiated by (–)sulpiride because, at a concentration of $1 \mu\text{M}$, SKF 38393 stimulates the D-1 dopamine receptor without having a significant effect on the D-2 receptor. However, the SKF 38393-

stimulated efflux of cyclic AMP was diminished by LY-141865 as a consequence of the interaction between LY-141865 and the second, pituitary-like D-2 dopamine receptor. Accordingly, this effect of LY-141865 was blocked by (–)sulpiride.

In the intermediate lobe of the rat pituitary gland, a β -adrenoceptor and a D-2 dopamine receptor occur on the same cell. Stimulation of the D-2 dopamine receptor decreases the

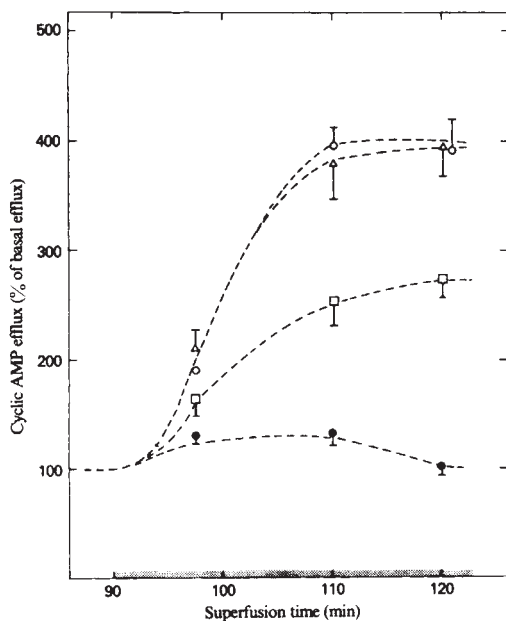


Fig. 3 Inhibitory effect of LY-141865 on SKF 38393-stimulated efflux of cyclic AMP is antagonized by (–)sulpiride. Experimental procedures were as described for Fig. 1a. Basal efflux of cyclic AMP was determined from three successive 10-min fractions collected between 60 and 90 min. Exposure to drugs (thickened abscissa) began at 90 min. Efflux of cyclic AMP was expressed as the mean % basal efflux ($\pm$ s.e.m.) of observations made on four superfusion chambers (as described for Fig. 1b). Tissue was superfused without drug (●), or with 10^{-6} M SKF 38393 (○), 10^{-6} M SKF 38393 and 10^{-6} M LY-141865 (□), or with 10^{-6} M SKF 38393, 10^{-6} M LY-141865 and 10^{-6} M (–)sulpiride (Δ).

formation of cyclic AMP stimulated by β -adrenergic agonists. By analogy with the situation in the intermediate lobe, the biphasic regulation of cyclic AMP formation by dopamine raises the possibility that both the D-1 and D-2 dopamine receptors are associated with the same cell in the neostriatum. However, it is possible that stimulation of the D-2 receptor decreases the responsiveness of the D-1 receptor by an indirect, trans-synaptic process. Although a β -adrenoceptor enhances cyclic AMP formation in the neostriatum^{15,28}, the failure of LY-141865 to affect significantly the (–)isoprenaline-stimulated efflux of cyclic AMP suggests that the D-2 dopamine receptor and the β -adrenoceptor are not functionally connected in the neostriatum. Although our data demonstrate a biochemical consequence of stimulation of the D-2 dopamine receptor in the neostriatum, the ultimate physiological consequences of stimulation of either the D-1 or the D-2 receptor in this area of the brain remain unknown.

Received 9 June; accepted 14 September 1981.

- Harmar, A. J. & Horn, A. S. *Molec. Pharmacol.* **13**, 512–520 (1977).
- Kebabian, J. W., Petzold, G. L. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 2145–2149 (1972).
- Brown, E. M., Carroll, R. J. & Aurbach, G. D. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4210–4213 (1977).
- Watling, K. J. & Dowling, J. E. *J. Neurochem.* **36**, 559–568 (1981).
- Kebabian, J. W. & Greengard, P. *Science* **174**, 1346–1349 (1971).
- Kebabian, J. W. & Calne, D. B. *Nature* **277**, 93–96 (1979).
- DeCamilli, P., Macconi, D. & Spada, A. *Nature* **278**, 252–254 (1979).
- Giannattasio, G., DeFerrari, M. E. & Spada, A. *Life Sci.* **28**, 1605–1612 (1981).
- Munemura, M., Eskay, R. L. & Kebabian, J. W. *Endocrinology* **106**, 1795–1803 (1980).
- Cote, T. E., Grewe, C. W. & Kebabian, J. W. *Endocrinology* **108**, 420–426 (1981).
- Munemura, M., Cote, T. E., Tsuruta, K., Eskay, R. L. & Kebabian, J. W. *Endocrinology* **107**, 1676–1683 (1980).
- Sibley, D. R. & Creese, I. *Endocrinology* **107**, 1405–1409 (1980).
- Creese, I., Schneider, R. & Snyder, S. H. *Eur. J. Pharmacol.* **46**, 377–381 (1977).
- Stefanini, E., DeVoto, P., Marchisio, A. M., Vernalone, F. & Collu, R. *Life Sci.* **26**, 583–587 (1980).
- Forn, J., Krueger, B. K. & Greengard, P. *Science* **186**, 1118–1120 (1974).
- Sala, G. B., Hayashi, K., Catt, K. J. & Dufau, M. L. *J. biol. Chem.* **254**, 3861–3865 (1979).
- Tait, J. F., Tait, S. A. S. & Bell, J. B. *Essays Biochem.* **16**, 99–174 (1980).
- Beavo, J. A. *et al. Molec. Pharmacol.* **6**, 597–603 (1970).
- Mulder, A. H., Van den Berg, W. B. & Stoof, J. C. *Brain Res.* **99**, 419–424 (1975).
- Setler, P., Sarau, H. M., Zerkle, C. L. & Saunders, H. L. *Eur. J. Pharmacol.* **50**, 419–430 (1978).

- Brown, E. M. *et al. Molec. Pharmacol.* **18**, 335–340 (1980).
- Scapagnini, U. *et al. in Sulpiride and Other Benzamides* (eds Spano, P. F., Trabucchi, M., Corsini, G. U. & Gessa, G. L.) 193–205 (Italian Brain Research Foundation, Milan, 1979).
- Spano, P. F., Stefanini, E., Trabucchi, M. & Fresia, P. *in Sulpiride and Other Benzamides* (eds Spano, P. F., Trabucchi, M., Corsini, G. U. & Gessa, G. L.) 11–31 (Italian Brain Research Foundation, Milan, 1979).
- Bach, N. J. *et al. J. med. Chem.* **23**, 481–491 (1980).
- Tsuruta, K. *et al. Nature* **292**, 463–465 (1981).
- MacLeod, R. M. *in Frontiers in Endocrinology* Vol. 4 (eds Martini, L. & Ganong, W. F.) 169–194 (Raven, New York, 1976).
- Caron, M. G. *et al. J. biol. Chem.* **253**, 2244–2253 (1978).
- Alexander, R. W., Davis, J. N. & Lefkowitz, R. J. *Nature* **258**, 437–440 (1975).
- Dowling, J. E. & Watling, K. J. *J. Neurochem.* **36**, 569–579 (1981).
- König, J. F. R. & Klippel, R. A. *in The Rat Brain: A Stereotaxic Atlas* (Krieger, New York, 1970).
- Harper, J. F. & Brooker, G. L. *J. Cyclic Nucleotide Res.* **1**, 207–218 (1975).

Single channel potassium currents of the anomalous rectifier

Yasuo Fukushima

Laboratory of Neurobiology, Institute of Brain Research, School of Medicine, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo, Japan

The recording of single channel currents has revealed the discrete events underlying the conductance changes for the Na and delayed K currents of the action potential, as well as those underlying acetylcholine-evoked currents^{1–5}. The existence of similar discrete events has been suggested by noise analysis of the current termed anomalous K rectifier^{6,7}, in which a K conductance is increased by hyperpolarization or by an increased extracellular concentration of K ions⁸. I have now used the patch current recording technique to investigate the anomalous rectifier of the tunicate egg cell. The statistics of pulse-like events in the patch current were found to be consistent with the inactivation kinetics of the anomalous rectifier. The extrapolated zero-current potential of the pulse-like events was approximately equal to that of the total anomalous K current. The single channel conductance was 5.2 pS in 50 mM K solution.

Tunicate egg cells were voltage-clamped with two micro-electrodes as described previously⁹. The bath saline contained 200 mM NaCl, 360 mM choline chloride and 10 mM MnCl₂, pH 7.0, and was maintained at 13.5–15.5 °C. A clean membrane surface was obtained by enzymatic digestion of the chorion. Patch current was recorded with a current-voltage converter which had a feedback resistor of 500 MΩ. The patch electrode was a conventional micropipette that had not been fire polished, and was filled with saline similar to that in the bath, except that 50, 100 or 200 mM K replaced an equal amount of choline. CaCl₂ (10 mM) and MgCl₂ (30 mM) were used as divalent salts; later 10 mM MnCl₂ was found to provide more stable recordings. The resistance of the patch pipette was 10–50 MΩ. When pushed against the egg surface, the seal resistance increased to >1 GΩ (usually 5–8 GΩ). The background noise level was 0.4–0.8 pA (peak-to-peak value, low-pass filtered at 660 Hz with a three-pole Butterworth filter). The rise time between 10 and 90% of the current step recorded by the patch electrode was 0.7–1.0 ms. Current records were digitized at 200 μs per point and stored by an 8-bit, 1K-word transient recorder (Kawasaki Electronics TM 1510).

After obtaining a large seal resistance, 150-ms hyperpolarizing pulses of up to –200 mV were applied to the whole egg cell, from a holding potential of –10 mV. Pulse-like changes in the patch current were observed which had a uniform amplitude at a given membrane potential (Fig. 1c). The pulse-like events in a patch usually disappeared within 20 command pulses. In some such cases, the amplitude of the pulse-like events was reduced in size during the last two or three command pulses and the traces for these were discarded. The current traces at –170 mV in

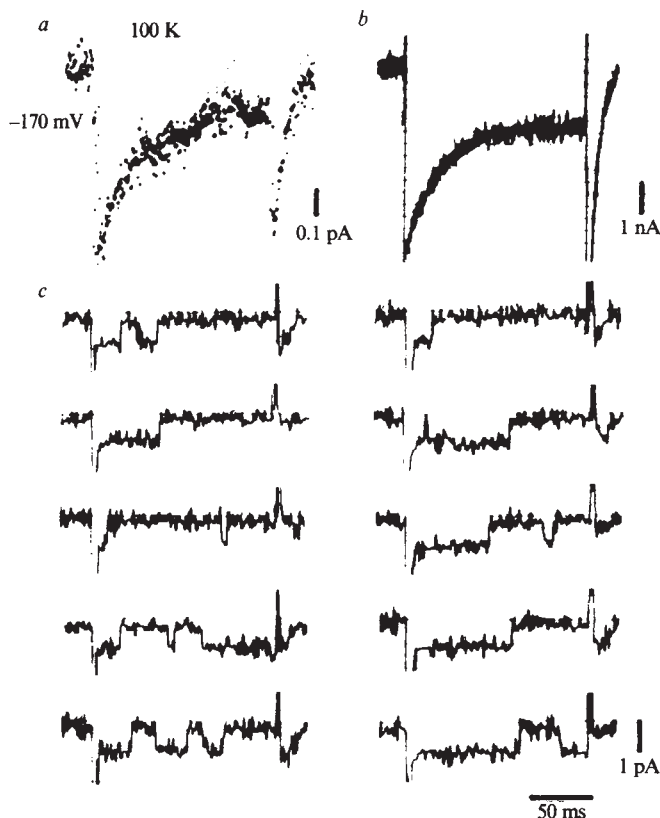
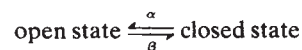


Fig. 1 *a*, Averaged time course of the pulse-like events in a patch. The whole egg cell was hyperpolarized to -170 mV during 45 command pulses. K concentration in the patch pipette was 100 mM, indicated as 100 K in this and other figures. *b*, Total membrane K current of the same egg cell in the same solution as in the patch pipette. The estimated number of channels was 7,200 per egg. *c*, Ten representative traces from the averaged 45 current traces. The inward current after the command pulse in *a*, *b* and *c* was Na current, which was evoked at the holding level of -10 mV after the hyperpolarizing command pulses relieved the inactivation of Na current.

100 mM K solution mostly started with inward deflections after the 3–6-ms capacitive transient and showed two to three pulse-like changes in the patch current during the command pulse lasting 150 ms. Multiples of the unitary amplitude were observed only exceptionally (Fig. 2*b*), indicating that, in most cases, a single channel existed under the patch pipette. Only the current traces from patches with one channel were analysed. Before or after patch recording the solution in the experimental bath was replaced by that contained in the patch pipette and a conventional voltage-clamp experiment was performed to examine the total anomalous K current. With a hyperpolarizing pulse of -170 mV the total anomalous K current showed an almost instantaneous rise and a subsequent slow exponential decay due to inactivation (Fig. 1*b*). When the pulse-like events in the patch current obtained at the same potential level were averaged, the time course was similar to that of the total anomalous K current (Fig. 1*a*). When the command pulse potential was less negative than -170 mV, the amplitude of the pulse-like events became smaller and the number of the events in a current trace decreased. Few or no events were observed at potentials more positive than -130 mV (Fig. 2*a*, -130 mV).

To analyse the kinetics of the pulse-like events in the patch current, the duration of the first open events was measured and it was denoted as the open time. The closed time was defined as the duration of the lower conductance period following the first open event (Fig. 3); unless it was terminated by another opening, the closed period was not counted. The distribution of the open or closed time showed a monotonic decay with duration; the distribution of the open time at -170 mV was roughly

exponential. The mean open time at -130 mV was markedly longer than at -170 mV, while the mean closed time at -130 mV was slightly shorter than at -170 mV. Assuming that the inactivation process can be described as a first-order transition between the open and closed states,



and that all channels are open immediately after the start of the command pulse, the probability density functions for the open time, $f_{\text{open}}(t)$, and the closed time, $f_{\text{closed}}(t)$, are

$$f_{\text{open}}(t) = \beta \exp(-\beta t)$$

$$f_{\text{closed}}(t) = \alpha \exp(-\alpha t)[1 - \exp\{-\beta(T-t)\}]$$

T is the duration of the command pulse, here equal to 150 ms.

The expressions are derived as follows. The probability that a channel remains open at a time t following its opening is $p = \exp(-\beta t)$, so the probability density function describing the probability that it closes exactly at t —that its lifetime is t —is

$$f_{\text{open}}(t) = -dp/dt = \beta \exp(-\beta t)$$

A similar expression would describe the probability density function for the closed times, except that the termination of the observation period T by the end of the command pulse reduces the probability of observing long closed times. The reduction depends on the distribution of the open times:

$$\begin{aligned} f_{\text{closed}}(t) &= \alpha \exp(-\alpha t) \int_0^{T-t} \beta \exp(-\beta t') dt' \\ &= \alpha \exp(-\alpha t)[1 - \exp\{-\beta(T-t)\}] \end{aligned}$$

The closing rate constants β , which were estimated from the total anomalous K currents, were 4.01 and 18.97 s^{-1} , and the opening rate constants α 12.78 and 9.46 s^{-1} , at -130 and -170 mV respectively. In Fig. 3, the probability density functions for single channels estimated from these rate constants are superimposed on the actual distribution by solid thick lines. The actual distributions were in good agreement with the estimated curves; deviations may be due to the small sample

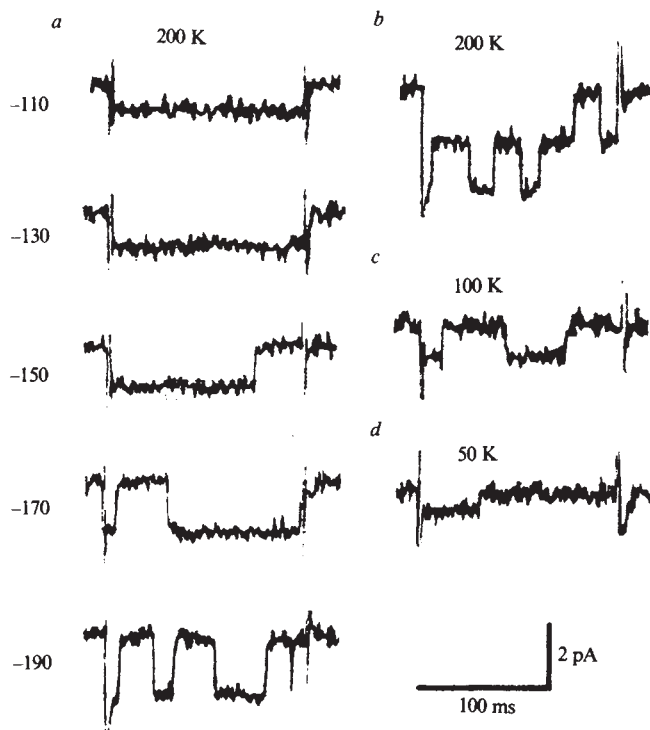


Fig. 2 *a*, Patch current in 200 mM K solution (200 K). The potential levels (in mV) during the command pulses are shown on the left. *b*, Current through a patch with two channels in 200 mM K solution. *c*, *d*, Patch current in 100 and 50 mM K solutions. The potential levels were -170 mV in *b*, *c* and *d*.

numbers. The asymmetric dependencies of the opening and closing rate constants on the membrane potential, which were observed from the decay time course of the total anomalous K current, were confirmed in the kinetics of a single channel. It was concluded that the statistics of single channel events in the patch current reflect the inactivation process of the anomalous K rectifier.

The amplitude of the pulse-like events in individual patches was constant at a given membrane potential, with some variations due to the background noise (Fig. 4a). The amplitude of the pulse-like events in a patch was linearly related to the membrane potential in the range -200 to -100 mV (Fig. 4b). Zero-current potentials of the pulse-like events were estimated by extrapolating the least-squares fits to the amplitudes for those patches from which current traces were recorded at least at three membrane potentials. The values of the estimated zero-current

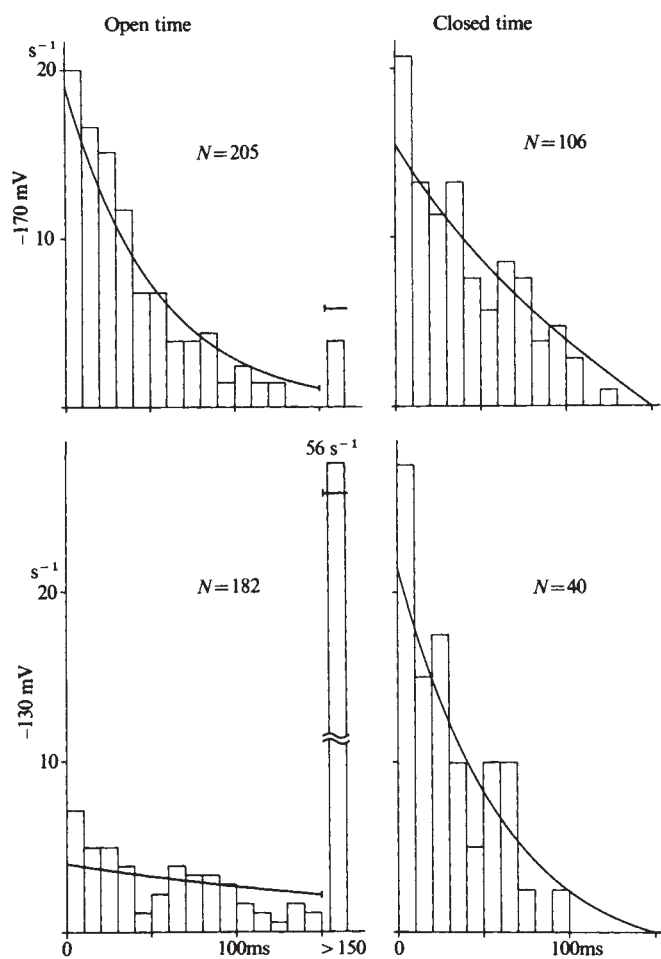


Fig. 3 The distribution of open and closed times, plotted as probability density. The patch current was recorded in 100 mM K solution. Only the first open and closed events of each record were measured, because later events were suspended by the end of the command pulse and accurate measurement of the durations was impossible. Cases in which the channel remained open during the 150-ms command pulse were plotted at the right edge in the distribution of the open times. Cases in which the channel failed to reopen during the command pulse are not included in the distribution of closed times. In the cases where the channel started in the closed state after the artefact, the open event was considered to have terminated during the artefact and was plotted in the first column in the distribution of the open time. The closed time in these cases was calculated by using an open time calculated from the duration of the artefact and the predicted distribution curve. N in each plot is the total sample number, which was normalized to 1 for comparison with the calculated probability density functions. The probability density function for the closed time was normalized by the integral from zero time to 150 ms.

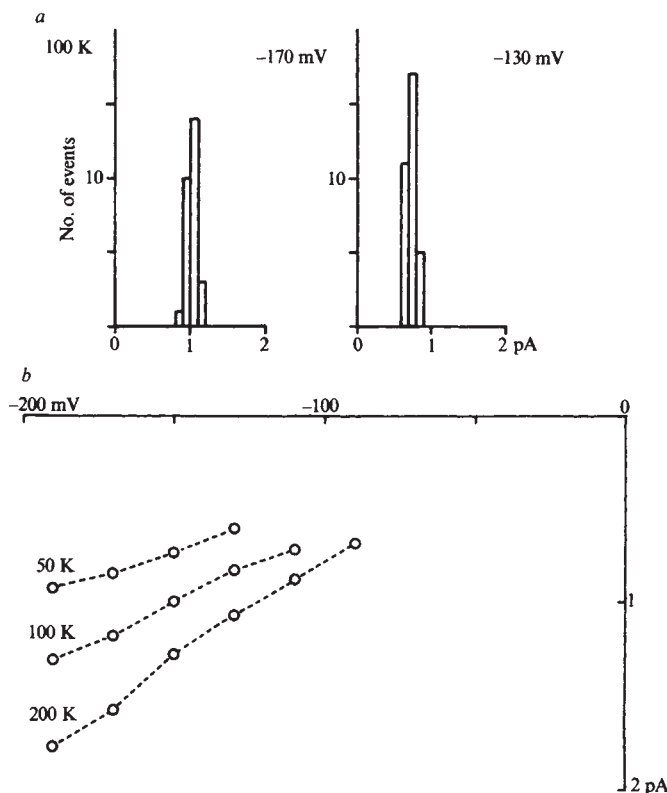


Fig. 4 *a*, Amplitude distribution of the pulse-like events in a single patch. The membrane potential was -170 and -130 mV in 100 mM K solution. *b*, Membrane potential and the amplitude of the pulse-like events in a patch. The K concentration in the patch pipette is indicated for each curve. The mean amplitude was plotted for several current traces recorded at a certain potential level.

potentials were -35 ± 7 , -30 ± 19 and -24 ± 14 mV (mean $\pm$ s.d.) in 50, 100 and 200 mM K solutions respectively. The zero-current potentials of the total anomalous K current were measured by the resting potentials in Na-free 50, 100 and 200 mM K solutions, giving values of -37 , -20 and -3 mV respectively. The difference in the zero-current potentials between those extrapolated from the amplitude of the pulse-like events in the patch current and those measured with the total anomalous K current may be due to a nonlinear current-voltage relationship for single channels.

At -170 mV in 50 mM K solution, the mean amplitude of the pulse-like events occurring in different patches was 0.69 ± 0.14 pA and the single channel conductance was 5.2 ± 1.0 pS, using a zero-current potential of the total anomalous K current. In 100 and 200 mM K solutions, the mean amplitude of the pulse-like events occurring in different patches showed more scattered distribution than that expected from the background noise. Because the statistics of the pulse-like events did not differ between patches, a heterogeneity of channel types is unlikely. Differences in the seal resistance bore no relation to the scattering of the amplitudes. The smaller amplitude of the events may be attributed to the channels under the patch pipette, as proposed in the case of the acetylcholine current¹. The number of channels in an egg was estimated as the ratio of the amplitude of the total anomalous K current in 50 mM K solution at the start of the command pulse to the single channel current of 0.69 pA at -170 mV, where the activation process was almost instantaneously completed and all channels were considered to be initially open. The estimated value was $9,300 \pm 6,000$ per egg ($0.041 \mu\text{m}^{-2}$).

Discrete conductance events have been reported in the delayed and Ca-dependent K current^{2,10,11}. My experiment shows that single conductance events also underlie the

anomalous K rectifier. The statistics of single channel events are concluded to be derived from the inactivation of the anomalous rectifier because the distribution of the open and closed times coincided with those estimated from the decay time course of the total anomalous K current. The single channel conductances for the anomalous rectifier estimated by the steady-state and ensemble noise analysis were 5.97 and 5.50 pS in 50 mM K solution^{6,7}. These values are in good agreement with the value of 5.2 pS in 50 mM K solution obtained in the present experiment. Recently the single channel current of the anomalous rectifier has been recorded in rat myotubes and conductance of similar order was obtained¹².

I thank Professor K. Takahashi for valuable advice at all stages of this work, Dr Y. Yajima for his suggestions in the mathematical analysis, and Professor A. Takeuchi and Dr H. Ohmori for criticism of the manuscript.

Received 1 June; accepted 22 September 1981.

1. Neher, F., Sakmann, B. & Steinbach, J. H. *Pflügers Arch. ges. Physiol.* **375**, 219–228 (1978).
2. Conti, F. & Neher, E. *Nature* **285**, 140–143 (1980).
3. Sigworth, F. J. & Neher, E. *Nature* **287**, 447–449 (1980).
4. Fukushima, Y. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1274–1277 (1981).
5. Horn, R., Patlak, J. & Stevens, C. F. *Nature* **291**, 426–427 (1981).
6. Ohmori, H. *J. Physiol., Lond.* **281**, 77–99 (1978).
7. Ohmori, H. *J. Physiol., Lond.* **311**, 289–305 (1981).
8. Hagiwara, S. & Takahashi, K. *J. Membrane Biol.* **18**, 61–80 (1974).
9. Okamoto, H., Takahashi, K. & Yoshii, M. *J. Physiol., Lond.* **254**, 607–638 (1976).
10. Lux, H. D., Neher, E. & Marty, A. *Pflügers Arch. ges. Physiol.* **389**, 293–295 (1981).
11. Marty, A. *Nature* **291**, 497–500 (1981).
12. Ohmori, H., Yoshida, S. & Hagiwara, S. *Proc. natn. Acad. Sci. U.S.A.* (in the press).

Gramicidin forms multi-state rectifying channels

David Busath & Gabor Szabo

Department of Physiology and Biophysics, University of Texas Medical Branch, Galveston, Texas 77550, USA

Gramicidin A is a pentadecapeptide of alternating L and D amino acids¹. In membranes it forms cation conductive channels² of molecular dimensions^{3,4} that in many respects resemble^{2–7} the channels of excitable cells^{8,9}. For this reason, and also because its structure is well established^{1,10,11}, the gramicidin channel is regarded as a useful model of ion channels in cell membranes. Although the properties of the gramicidin channel have been studied extensively¹², it has generally been described as having a single state associated with a sharply defined conductance that remains unaltered during the lifetime of the channel¹³. We report here that, in fact, gramicidin A can assume other, less conductive 'miniature' (mini) states evidenced by our observations of spontaneous transitions in the conductance of single open channels and the observation of a significant number of weakly conducting channels. Current–voltage (*I*–*V*) relationships for different channel states differ significantly and, for minis, are often asymmetrical. Our results indicate that the gramicidin channel has a wide variety of stable conformational states that give rise to channels with different electrical properties. Because transitions between states occur relatively infrequently, these conformational states must be separated by relatively large energies of interconversion. Similar transitions, poised by the electric field or an agonist molecule, may underlie the function of gated channels in cell membranes.

Gramicidin A was purified by HPLC from commercially available gramicidin D (a mixture¹⁴ of gramicidins A, B and C). Membrane currents in the presence of small amounts of gramicidin A were amplified and sampled digitally for computer-assisted analysis. For channel conductance measurement (Fig. 1), the electrical potential across the membrane was fixed so that the ionic current through single channels could be monitored directly. For single-channel *I*–*V* measurements (Fig. 2), the

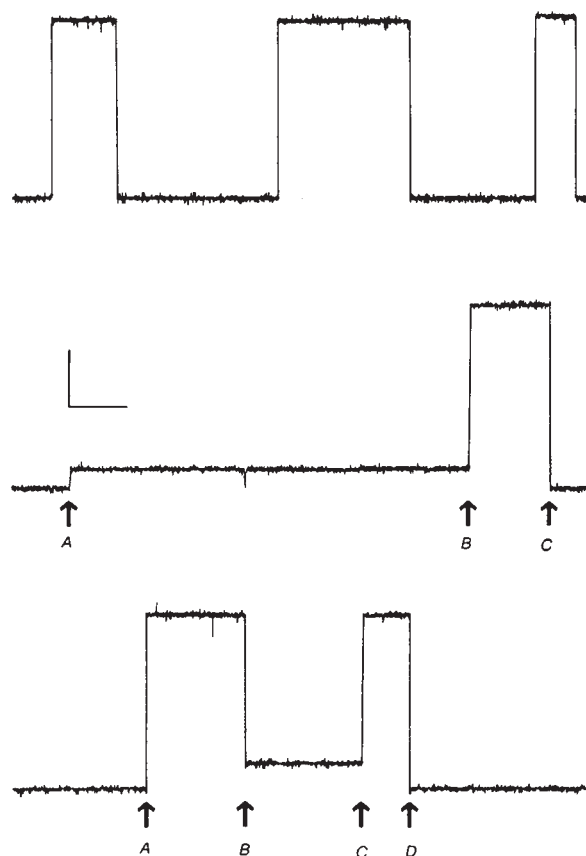


Fig. 1 Time course of the ionic current through single gramicidin A channels. HPLC-purified gramicidin A was added to a bilayer formed from a dispersion of 1-monoolein (50 mg, Nu-Check) in squalene (1 ml, Sigma, purified through silica) and held at 0.1 V. Vertical calibration bar, 1.5 pA. Horizontal bar, 35 s (upper), 10 s (middle) and 15 s (lower). The signal was bandlimited to 30 Hz and sampled 30 times per s. 1M KCL, pH ~5.6, $23 \pm 1^\circ\text{C}$.

membrane potential was varied at a constant rate. Direct recording of single-channel *I*–*V* curves was made possible by reducing the membrane area ($1.5 \times 10^{-5} \text{ cm}^2$) so that capacitive current would not obscure the single-channel current.

Channels formed by purified gramicidin A turn on abruptly, that is, with a time course ($<200 \mu\text{s}$) that cannot be experimentally resolved. The channels typically remain conductive in a state that has a uniform conductance, $47.5 \pm 0.5 \text{ pS} \pm \text{s.d.}$, in the conditions of Fig. 1, and cease conducting abruptly and completely¹⁵. Three examples of these 'typical' channels are shown in the upper trace of Fig. 1. By recording the occurrence of many single channels, we have been able to observe a small but well defined fraction of channels that have novel, atypical properties. The middle trace of Fig. 1 shows, for example, a weakly conducting 'mini' channel (4.6 pS) turning on at point A. At point B the mini channel abruptly becomes more conductive, with a conductance (47.9 pS) characteristic of typical channels, and at point C this channel ceases to conduct. The lower trace of Fig. 1 illustrates that channel conductance may also spontaneously decrease. On this trace a mini channel of high conductance (45.2 pS) turns on at point A, at point B the conductance decreases abruptly and spontaneously to 6 pS and after returning to 45.2 pS at C, the channel turns off at D. We considered whether such traces could merely be the coincidental occurrence of two transitions (such as a turn-on and a turn-off) in the same sampling interval. By using very thin monoolein/squalene bilayers¹⁵ (and thus maximizing the single-channel lifetimes), applying only small amounts of gramicidin and sampling rapidly compared with the transition rate, we were able to minimize the possibility of two transitions occurring in

the same sampling interval. Twenty-six spontaneous state changes were observed in a contiguous, prospectively studied set of data lasting 8.25 h (8.9×10^5 sampling intervals) and containing 424 channels. In our conditions of measurement, the probability that all 26 events resulted from coincidental transitions is <1 in 10^{41} . Therefore we must be observing conductance changes in the same channel. Clearly the same gramicidin channel can assume multiple conductance states.

The diminished conductance of the mini state must arise from a spontaneously increased barrier to ion transport or increased binding of ions in the pore. If such a change occurred at the centre of the pore, the I - V relationship should be symmetrical, as it is for the main channel (Fig. 2a). However, if this change occurred near either end of the pore, an asymmetrical I - V relationship^{16,17} would be expected. We have found that I - V relationships of mini channels are frequently asymmetrical, as illustrated in Fig. 2. Curve *a* shows the I - V relationship for a single typical channel recorded for 19 sweeps of the I - V curve. Before turning off, the channel converted to a mini channel and the I - V relationship was recorded for six sweeps of the membrane potential (curve *b*). Curve *c* shows 11 sweeps of the I - V curve for a single mini channel which did not change state during its lifetime. The broken line is drawn as a reference to demonstrate the asymmetry of the mini channels compared with the symmetry and linearity of the I - V curve of typical channels (curve *a*).

Although the spontaneous changes between conductance states are infrequent, the occurrence of mini channels which do not change states is rather frequent. This remains the case after extensive purification^{6,18-20} and *de novo* synthesis^{20,21} of gramicidin A. Figure 3 shows the relative frequency of minis in the conductance transition histogram. Of the 354 channels that occurred during the experiment, 199 (56%) were mini channels. In other experiments the fraction of minis ranged from 22% to

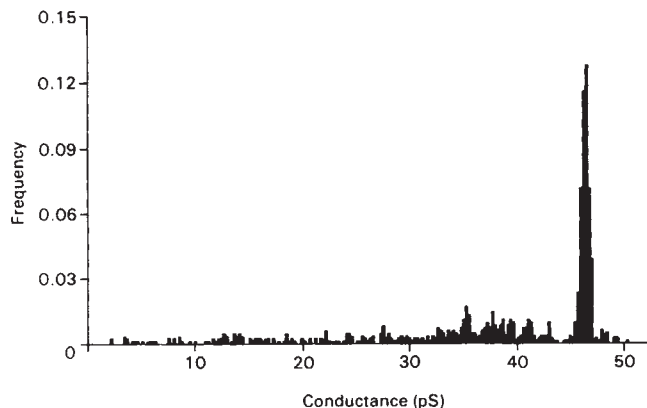


Fig. 3 Histogram of single-channel conductances using HPLC-purified gramicidin A. The sharp peak has a conductance of 46.0 ± 0.3 (s.d.) pS and contains 44% of the transitions. The remainder are minis which we define as those with a conductance more than two standard deviations below the mean of the main gaussian peak. Glyceryl-1-monoolein (50 mg)/hexadecane (1 ml, Aldrich Gold Label), 1 M KCl, pH ~ 5.6 , 0.1 V, $23 \pm 1^\circ\text{C}$.

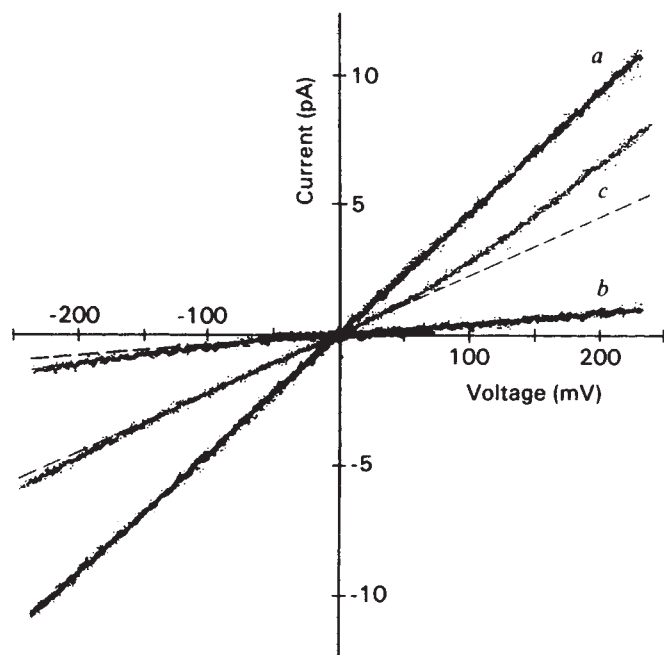


Fig. 2 Single-channel I - V relationships for a typical channel, (*a*, 45.3 pS) and a mini channel into which it converted (*c*). Also shown is a mini channel which did not change states during its lifetime (*b*). Purified gramicidin A was added to a glyceryl-1-monoolein/squalene (50 mg ml^{-1}) bilayer and a triangular voltage wave (0.5 Hz) applied. Current and voltage were filtered at 100 Hz and sampled simultaneously 300 times per s. Baseline-cycle currents were subtracted digitally from currents in cycles where a single channel was conducting and the result plotted against the corresponding measured voltage. The broken reference lines have slopes of 22.7 pS (*c*) and 3.8 pS (*b*). 1 M KCl, pH ~ 5.6 , $23 \pm 1^\circ\text{C}$.

50%, varying within the expectations based on Poisson statistics. The conductance of minis has a broad distribution over 0–45 pS without any sharply defined features. This implies a profusion of different states whose probabilities of occurrence are roughly the same. The low frequency of spontaneous changes between conductance states ($<0.0013 \text{ s}^{-1}$) indicates that the different states correspond to stable conformations that are separated by large free-energy barriers. A rudimentary calculation, based on absolute rate theory, gives a barrier height of $\geq 21 \text{ kcal mol}^{-1}$.

We found that the rate of channel opening for minis was directly proportional to the rate of channel openings for typical channels over a 1,000-fold range. This indicates that the turn-on frequencies for mini channels and typical channels depend in the same way on membrane gramicidin concentration. As channel turn-on frequency is proportional to the square of gramicidin concentration^{16,22,23}, both mini and typical channels must result from the association of two monomers. Thus minis cannot be membrane-spanning monomers, laterally associated trimers or tetramers, or trimers or tetramers in series⁶.

We also found that the mean lifetime of minis does not differ from that of typical channels in either GMO-squalene (40–60 s) or GMO-hexadecane (2–3 s) bilayers. The similar lifetimes of minis and typical channels indicate that (1) minis are not normal channels in thicker patches of membrane⁶ and (2) the mini channel structure could not differ greatly from the β -helical^{4,10} head-to-head^{24,25} dimer structures of typical channels. Were minis of a radically different conformation^{26,27} (such as a parallel helix, as proposed by Veatch²⁶), their lifetime should also be different from that of the typical channels because a different number of hydrogen bonds must be broken to dissociate such a dimer. Similar arguments rule out other variations of the helical pitch⁴ of dimers at the head-to-head junction which would require a different activation energy for dimer dissociation.

Three classes of mechanism for the formation of mini channels would seem consistent with our observations. First, minis could result from the influence of an external factor on the channel, for example a dipolar or charged moiety^{16,17} binding tightly to the external surface of one end of the pore and thereby reducing its ion conductance. Second, minis could result from defects in the β -helical structure of the peptide backbone away from the head-to-head junction, as suggested by structural studies¹¹ indicating that the presence of an ion in the channel may alter the pitch of the backbone helix. Finally, as proposed by Urry *et al.*²⁸, changes in the positions of the side chains, particularly the large tryptophan and leucine R groups, could decrease channel conductance asymmetrically. These side-chain

effects could be mediated by changes in the equilibrium positions of the backbone carbonyls, by changes in the local electric field or the electron-donating capacity of carbonyls, or by changes in the libration energy^{10,28} of the protein backbone. If conductance states are determined by side-chain positions, then it is reasonable to speculate that dipolar or reactive residues could couple these states to the electrical potential and/or the presence of agonist molecule, resulting in electrically or chemically gateable channels.

We thank Drs R. E. Koeppe and Dan Urry for gifts of purified gramicidin A, and Drs Wayne Giles, Ron Waldbillig and Don McBride for their comments and suggestions. This work was supported by DHHS grants GM 26897 and HL 24820.

Received 24 July; accepted 22 September 1981.

1. Sarges, R. & Witkop, B. *J. Am. chem. Soc.* **86**, 1862–1863 (1964).
2. Myers, V. B. & Haydon, D. A. *Biochim. biophys. Acta* **274**, 313–322 (1972).
3. Urry, D. W. *Proc. natn. Acad. Sci. U.S.A.* **68**, 672–676 (1971).
4. Urry, D. W., Goodall, M. C., Glickson, J. D. & Mayers, F. D. *Proc. natn. Acad. Sci. U.S.A.* **68**, 1907–1911 (1971).
5. Neher, E., Sandblom, J. & Eisenman, G. *J. Membrane Biol.* **40**, 97–116 (1978).
6. Hladky, S. B. & Haydon, D. A. *Biochim. biophys. Acta* **274**, 294–312 (1972).
7. Shagina, L., Grinfeldt, A. E. & Lev, A. *Nature* **273**, 243–245 (1978).
8. Begenisich, T. B. & Cahalan, M. D. *J. Physiol., Lond.* **307**, 217–242 (1980).
9. Hodgkin, A. L. & Keynes, R. D. *J. Physiol., Lond.* **128**, 61–88 (1955).
10. Urry, D. W. in *Conformation of Biological Molecules and Polymers—The Jerusalem Symp. on Quantum Chemistry and Biochemistry* Vol. 5, 723–736 (Jerusalem Academy of Sciences, 1973).
11. Koeppe, R. E., Hodgson, K. O. & Stryer, L. *J. molec. Biol.* **121**, 41–54 (1978).
12. Bamberg, E. *et al. Fedn Proc.* **37**, 2633–2638 (1978).
13. Haydon, D. A. & Hladky, S. B. *Q. Rev. Biophys.* **5**, 187–282 (1972).
14. Gross, E. & Witkop, B. *Biochemistry* **4**, 2495–2501 (1965).
15. White, S. *Biophys. J.* **23**, 337–347 (1978).
16. Apell, H.-J., Bamberg, E., Alpes, H. & Luger, P. *J. Membrane Biol.* **31**, 171–181 (1977).
17. Frohlich, O. *J. Membrane Biol.* **48**, 365–383 (1979).
18. Hladky, S. B. & Haydon, D. A. *Nature* **225**, 451–453 (1970).
19. Bamberg, E. & Luger, P. *Biochim. biophys. Acta* **367**, 127–133 (1974).
20. Prasad, K. U., Trapani, T. L., Busath, D., Szabo, G. & Urry, D. W. *Int. J. Peptide Protein Res.* **18** (in the press).
21. Bamberg, E., Noda, K., Gross, E. & Luger, P. *Biochim. biophys. Acta* **419**, 223–228 (1976).
22. Bamberg, E. & Luger, P. *J. Membrane Biol.* **11**, 177–194 (1973).
23. Veatch, W. R., Mathies, R., Eisenberg, M. & Stryer, L. *J. molec. Biol.* **99**, 75–92 (1975).
24. Bamberg, E., Appel, H.-J. & Alpes, H. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2402–2406 (1977).
25. Bamberg, E. *et al. J. Membrane Biol.* **50**, 257–270 (1979).
26. Veatch, W. R., Fossel, E. T. & Blout, E. R. *Biochemistry* **13**, 5249–5256 (1974).
27. Ochinnikov, Y. A. *Eur. J. Biochem.* **94**, 321–336 (1979).
28. Urry, D. W., Venkatachalam, C. M. & Prasad, K. U. *J. Quantum Chem., Quantum Biol. Symp.* **8** (in the press).

Lateral diffusion of lipids in sarcoplasmic reticulum membranes is area limited

Peter Laggner*

Institut für Röntgenfeinstrukturforschung der Österreichischen Akademie der Wissenschaften und des Forschungszentrums Graz, Steyrergasse 17, A-8010 Graz, Austria

Scandella *et al.*¹ used the technique of electron spin resonance to investigate lateral lipid diffusion in the membrane of sarcoplasmic reticulum from skeletal muscle. Their results suggested that lipids are free to diffuse within the plane of the membrane at rates comparable with those in isolated phospholipid bilayer systems of similar composition^{2–4}. However, their study¹ was carried out at temperatures between 50 and 70 °C. I have therefore now carried out similar studies, but at physiological temperatures and using fatty acid analogue spin labels. My results show absence of spin-exchange interaction at spin label concentrations below 2–4 mol %. This suggests the existence of diffusion barriers in the plane of the membrane which separate the lipids into clusters containing, on average, about 50 phospholipid molecules.

* Present address: European Molecular Biology Laboratory, Hamburg Outstation, c/o DESY, Notkestrasse 85, D-2 Hamburg, FRG.

Table 1 Size of lipid bilayer clusters and diffusion coefficients

<i>T</i> (°C)	<i>n</i>	<i>D</i> ($\times 10^8$ cm ² s ⁻¹)
SR membranes		
20	36	1.5
30	44	1.7
40	50	2.2
Sonicated membranes		
20	—	1.6
SR lipids		
20	—	1.7
40	—	2.3

n, Number of phospholipid molecules per bilayer cluster calculated from the concentrations at the break points in Fig. 1*b*. SR, sarcoplasmic reticulum.

Electron spin resonance (ESR) measurements were carried out between 20 and 40 °C using a stearic acid analogue spin label, 2-(3-carboxypropyl)-4,4-dimethyl-2-tridecyl-3-oxazolidinoxyl ('5NS'; Syva). This type of spin label has been useful in estimating lateral diffusion rates in *Escherichia coli* membranes⁵. Its advantage over the phosphatidylcholine analogue used by Scandella *et al.*¹ lies in the relative ease of its introduction into the membrane system, thus avoiding the poorly controllable process of fusion between phosphatidylcholine vesicles and the membrane, which is likely to cause a major structural perturbation in the membrane region where it occurs. The obvious disadvantage of the stearic acid label lies in its structural difference from endogenous lipids.

The method is based on the notion that the collision of spin labels introduced into the fluid lipid phase of the membrane is controlled by diffusion. Since colliding free nitroxide radicals will undergo spin-exchange interaction causing broadening of the ESR lines^{2,3,6}, the collision frequency can be monitored by linewidth measurement. Although quantitative evaluation in terms of diffusion coefficients must rely on certain specific assumptions about the diffusion model^{4,6}, the observation of concentration-dependent line broadening unambiguously indicates the occurrence of label diffusion.

Figure 1 shows that, within the temperature range used, no concentration-dependent line broadening of the ESR spectrum was observed below certain critical spin label concentrations, indicating that no collisions take place up to these concentrations. Above these concentrations, the linewidth increased, as expected for diffusion-controlled spin label collisions. The same behaviour was found with a different stearic acid analogue spin label, 2-(10-carboxydecyl)-2-hexyl-4,4-dimethyl-3-oxazolidinoxyl (12NS; Syva), while the methyl ester of 12NS showed the same linearly ascending slope at even the lowest concentrations (0.5 mol %).

By adopting the diffusion model and numerical factors described in detail by Sackmann *et al.*⁵, the diffusion constants calculated from the slopes above the critical concentrations (Table 1) correspond satisfactorily to the values determined previously for phosphatidylcholine spin labels¹. This consistency strengthens the reliability of results obtained by stearic acid spin labels. Parallel experiments with extracted membrane phospholipids in aqueous dispersion have verified the fact that at these low concentration increments line broadening could indeed be measured. The slopes observed with the isolated lipid samples were only slightly higher than with the membrane system, confirming that, once diffusion control of collisions prevails, the diffusion rates are very similar in the two systems.

Quantitative evaluation of the hyperfine peaks in terms of the order parameter *S*₃ and the isotropic coupling constant *a*_N (refs 7, 8) have shown constant values of 0.68 and 14.6 respectively, up to label concentrations of 10 mol %. This eliminates the possibility that the observed absence of collisions is due to binding of the first added spin labels by protein. Therefore, to interpret these effects, I postulate the existence of diffusion barriers in the plane of the membrane separating individual clusters of lipids. Table 1 lists the size of these clusters estimated

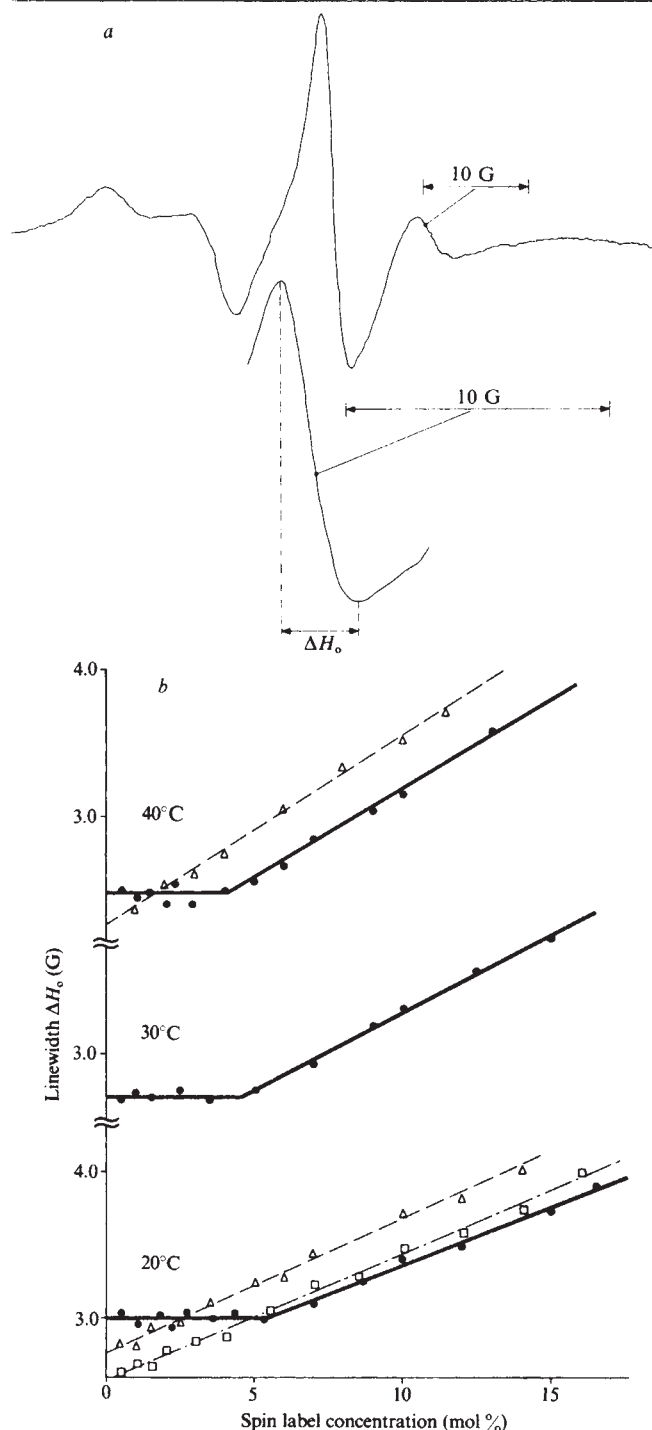


Fig. 1 *a*, Typical ESR spectrum from stearic acid analogue 5NS in sarcoplasmic reticulum membranes at label concentrations of 1 mol %, relative to total phospholipid (taken as molecular weight 750). Spectrometer settings were as follows. Field set 3,359 G; microwave power 10 mW; receiver gain 5×10^3 ; modulation amplitude 0.5 G; recorder time constant 0.5 s; scan time 4 min. The central line was recorded in a scan range of 40 G to facilitate precise linewidth measurement. A Varian E 104-A X-band spectrometer with variable temperature control unit was used. Sarcoplasmic reticulum membranes were prepared according to previously documented methods^{10,11} and studied within 48 h of preparation. Spin labels were introduced by gently shaking membrane suspensions, typically containing $\sim 20 \text{ mg ml}^{-1}$ phospholipids, under N_2 atmosphere in a glass tube containing the appropriate amounts of dry label deposited on the wall. *b*, Linewidths of the central ESR line of 5NS in sarcoplasmic reticulum membranes (●), isolated membrane phospholipid suspensions (Δ) and a sample of sarcoplasmic reticulum that had undergone 1 min ultrasonication (□) using the 1-cm probe tip of a Braun Labsonic sonicator.

from the critical onset concentrations for line broadening. For simplicity, this calculation involves the assumption that the clusters are phospholipid bilayers, that is, contain twice the number of molecules as deduced from the critical concentrations.

At physiological temperatures, the clusters would contain, on average, 50 phospholipid molecules, which is about half the number of phospholipids per molecule of $(\text{Ca}^{2+} + \text{Mg}^{2+})\text{ATPase}$, the major integral protein constituent of the membrane. It could be speculated that this reflects the fraction of phospholipids that are not involved in forming an 'annulus'⁹ around the ATPase which, by definition, would diffuse at a considerably slower rate.

The above estimations imply uniform contributions to the measured linewidths by all spin labels. This is supported, at concentrations $< 5 \text{ mol } \%$, by the absence of any free-label signal, indicating that all spin labels are incorporated into the membranes. Even above this concentration, the area under the subsequently gradually occurring free-label signal is only a minor fraction of the total and hence safely negligible for the present considerations. On the other hand, it cannot be excluded that a small number of spin label molecules are immobilized by binding to protein below the detection level. However, this also is negligible in relation to the concentrations relevant here.

Although it is premature to speculate on the detailed nature of the diffusion barriers, the finding that neither lipids alone nor ultrasonically perturbed membranes show these effects suggests that protein-protein interaction is probably involved. The observation that only the free stearic acid analogues and not the methyl ester are sensitive to this effect furthermore suggests that ionic interactions might be relevant in defining the limits of lipid diffusion. Note that the present results allow direct conclusions only to negatively charged lipids. Future experiments must show whether this also holds true for positive or zwitterionic lipids.

I thank Mrs U. Rakusch for technical assistance and Dr J. Suko for supplying several membrane samples. This work was supported under grant 3524 by the Österreichischer Fonds zur Förderung der Wissenschaftlichen Forschung.

Received 15 July; accepted 9 October 1981.

- Scandella, C. J., Devaux, P. & McConnell, H. M. *Proc. natn. Acad. Sci. U.S.A.* **69**, 2056 (1972).
- Devaux, P. & McConnell, H. M. *J. Am. chem. Soc.* **94**, 4475 (1972).
- Traeble, H. & Sackman, E. *J. Am. chem. Soc.* **94**, 4499 (1972).
- Cherry, R. J. in *Biological Membranes* Vol. 3 (eds Chapman, D. & Wallach, D. F. H.) 47-102 (Academic, London, 1976).
- Sackmann, E., Traeble, H., Galla, H. J. & Overath, P. *Biochemistry* **12**, 5360 (1973).
- Molin, Yu. N., Salikhov, K. M. & Zamaraev, K. I. *Spin Exchange, Principles and Applications in Chemistry and Biology* (Springer, Berlin, 1980).
- Seelig, J. in *Spin Labeling, Theory and Applications* (ed Berliner, L. J.) 373-409 (Academic, New York, 1967).
- Hubbell, W. L. & McConnell, H. M. *J. Am. chem. Soc.* **93**, 314 (1971).
- Hesketh, T. R. *et al. Biochemistry* **15**, 4145 (1976).
- Martonosi, A., Donley, J. & Halpin, R. A. *J. biol. Chem.* **243**, 61 (1968).
- Davis, D. G. & Inesi, G. *Biochim. biophys. Acta* **241**, 1 (1971).

Variant insertion element IS1 generates 8-base pair duplications of the target sequence

Shigeru Iida, Roberto Marcoli* & Thomas A. Bickle

Microbiology Department, Biozentrum, University of Basle, Klingelberstrasse 70, CH-4056 Basle, Switzerland

Transposition in both eukaryotes and prokaryotes is characterized by the generation of short, direct duplications of the target DNA sequence at the integration site of the transposable element¹. For example, the bacterial insertion sequence IS1, or

* Present address: ISREC, Chemin des Boveresses, CH-1066 Epalinges s/Lausanne, Switzerland.

the IS1-flanked transposon Tn9, generates 9-base pair (bp) repeats during either transposition or cointegration²⁻⁸. All the transposable elements so far sequenced are characterized by the presence at their ends, of a perfect, or near-perfect, inverted repeat sequence, which is ~30 bp long¹ and is thought to be important for recognition by enzymes during transposition. While studying IS1-mediated cointegration⁹⁻¹², we found an IS1 carrying a single base pair change in the terminal inverted repeat sequence. We now report that it generates 8-bp duplications of the target sequence, a finding which confirms the importance of the inverted repeat for transposition and which, because one end only of the IS1 is changed, has implications for the mechanism of transposition by IS1.

We have characterized four pBR325 derivatives carrying IS1, all of which resulted from segregation of cointegrates between pBR325 and plasmid P1, which contains an IS1 element¹¹⁻¹³ (Fig. 1). The sites of insertion and the orientation of the IS1 in these different derivatives have been described previously^{11,12}, and the sequencing strategy used is shown in Fig. 2.

Figure 3 shows the results of the DNA sequence analysis around the junctions of IS1 and pBR325. In each plasmid, the IS1 had integrated at a different position within a 300-bp region of pBR325 between coordinates 4,134 and 4,411 (ref. 14). Two of the plasmids (pSHI211 and pSHI212) carried the 9-bp duplications of target sequence previously reported for IS1²⁻⁸ but the other two (pSHI213 and pSHI214) had only 8-bp

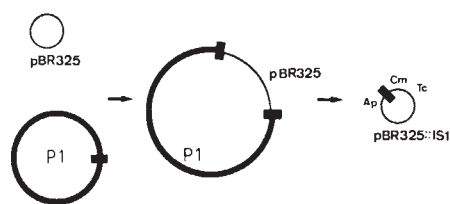


Fig. 1 The formation of transpositional cointegrates. Phage P1 was induced in P1-lysogenic Rec⁺ bacteria containing pBR325, which codes for resistance to ampicillin (Ap), chloramphenicol (Cm) and tetracycline (Tc)^{14,31}, and cointegrates were isolated as specialized P1 transducing phages. Their structure is shown in the centre of the figure: the P1 genome (thick line) has integrated into the pBR325 (thin line) and is flanked by direct repeats of the IS1 element resident on the P1 genome (bar). Subsequent resolution of these cointegrates in Rec⁺ cells yields a pBR325 derivative carrying a single copy of IS1^{11,12}.

duplications. The IS element of the latter two has a G to T transversion at position 757, 12 bp from the end of the element, and no other alteration within 50 bp of the end when compared with the canonical sequence^{3,15,16}. The same alteration has also been found in one of the flanking IS1s of Tn9⁸. There seems to be a good correlation, in our experimental conditions, between the alteration in the inverted repeat of IS1 and the ability of the element to generate 8-bp repeats.

The transposition characteristics of the variant IS1 seem very similar to those of IS1 itself. For example, in both cases the preferred sites for IS1 integration are A+T rich^{5-8,17-19} and both elements show a strong preference for inserting in sites that have a G-C base pair at each end of the target sequence to be duplicated⁸. The variant IS1 stimulates the formation of deletions, as does the wild type, and transposes with a slightly higher frequency^{11,12}.

IS1-mediated transpositional cointegrations have been observed in various experimental systems^{7,9,11,12,20}. As only one of the two parental plasmids originally contained an IS1 and the resulting cointegrates carry two IS1s at the junction of the replicons (Fig. 1), one can assume that the IS1 duplicates during

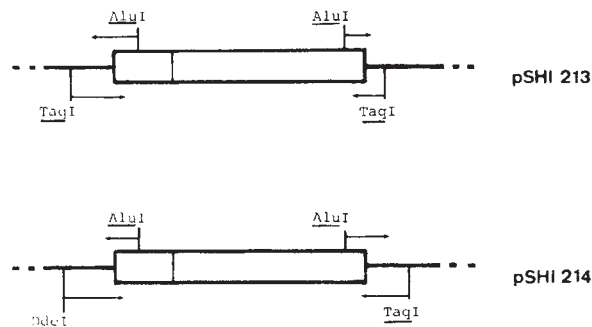


Fig. 2 Sequencing strategy for the variant IS1 ends. IS1 is drawn as a box and the flanking DNA sequences as a horizontal line; the vertical line within the box shows the orientation of IS1 and indicates the position of the single PstI site. Plasmids pSHI213 and pSHI214 had IS1 integrated into pBR325 at locations 4,404–4,411 and 4,134–4,141 respectively, and were sequenced as indicated by arrows. The other two plasmids, pSHI211 and pSHI212, had IS1 integrated at positions 4,202–4,210 and 4,248–4,256 respectively, and the ends were sequenced using the AluI sites of IS1. DNA sequencing was done using the method of Maxam and Gilbert³².

the process. If this is the case, the IS1 elements carried by the four pBR325 derivatives that we have analysed ought to have the same DNA sequence as the IS1 of P1. However, we find that two of the four carry an alteration. The possibility that the P1 parental phage carried heterogeneous IS1s is less likely because the cointegrates were isolated from fresh lysogens derived from a single plaque. Alternatively, a chromosomal IS1 may first have transposed into pBR325 and the cointegrates then have arisen by reciprocal recombination between the IS1 acquired by pBR325 and the IS1 of P1; such a pathway has recently been demonstrated in Rec⁺ cells¹². The most likely explanation is that two of the four cointegrates studied were formed by this pathway while the other two were formed by the IS1 duplication process. The same G to A transversion has been found in the flanking IS1 of Tn9 carried on λ cam1⁸ and this Tn9 originally derived from P1Cm0, which has Tn9 inserted within the IS1 of P1 (refs 13, 21–23). Therefore, the variant IS1 probably comes from the P1 genome.

The recently proposed models for transposition fall into two groups differing in the mechanism of transposition initiation. In group I models, only one end of the IS element is ligated to the target sequence at the initiation step²⁴⁻²⁶, whereas group II models predict that both ends of the IS element are ligated^{27,28}. A cointegrate is a necessary intermediate in group II models whereas it is a by-product in group I models. Evidence supporting transposition of IS1 by a group I mechanism comes from the study of IS1-mediated cointegrates. Their formation seems to be less frequent than transposition, but once formed they are

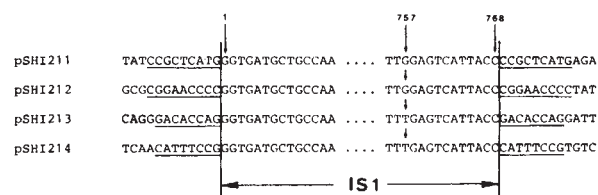


Fig. 3 DNA sequence of the ends of IS1 and flanking sequences. The ends of the IS1 are delineated by vertical lines. Duplicated plasmid sequences are underlined. Arrows indicate the position of the variation in the terminal inverted repeat of IS1 carried by pSHI213 and pSHI214. IS1 coordinates are from ref. 15.

stable, even in recombination-proficient cells^{7,10-12}, which argues that they are not intermediates in transposition. This contrasts with some other transposons, such as Tn3, where the cointegrates are unstable²⁸⁻³⁰; such transposons probably use a group II mechanism. Our observation that an alteration in one end only of IS1 changes its transposition characteristics also favours a group I mechanism. In this respect, it is interesting that the Tn9 with the same base alteration has been shown to generate 9-bp repeats^{4,8}. It is unlikely that the repeat generated depends on whether the process is transposition or cointe-

gration. The end of the IS element chosen to initiate transposition may be dictated not only by the sequence of the IS element itself, but also by sequences in the donor DNA molecule, perhaps the recognition sequences in the donor that were used when it previously acquired the IS element.

We thank W. Arber for advice and encouragement and D. J. Galas, N. D. F. Grindley, J. H. Miller, E. Ohtsubo and P. Starlinger for discussion and comments on an earlier version of this paper. The work was supported by grants from the Swiss NSF.

Received 12 August; accepted 2 October 1981.

- Calos, M. P. & Miller, J. H. *Cell* **20**, 579-595 (1980).
- Calos, M. P., Johnsrud, L. & Miller, J. H. *Cell* **13**, 411-418 (1978).
- Grindley, N. D. F. *Cell* **13**, 419-426 (1978).
- Johnsrud, L., Calos, M. P. & Miller, J. H. *Cell* **15**, 1209-1219 (1978).
- Devos, R., Contreras, R., van Emmelo, J. & Fiers, W. *J. molec. Biol.* **128**, 621-632 (1979).
- Kühn, S., Fritz, H. J. & Starlinger, P. *Molec. gen. Genet.* **167**, 235-241 (1979).
- Ohtsubo, E., Zenilman, M. & Ohtsubo, H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 750-754 (1980).
- Galas, D. J., Calos, M. P. & Miller, J. H. *J. molec. Biol.* **144**, 19-41 (1980).
- Iida, S. & Arber, W. *Molec. gen. Genet.* **177**, 261-270 (1980).
- Iida, S. *Plasmid* **3**, 278-290 (1980).
- Iida, S., Meyer, J. & Arber, W. *Cold Spring Harb. Symp. quant. Biol.* **45**, 27-43 (1981).
- Iida, S., Meyer, J. & Arber, W. *Molec. gen. Genet.* (in the press).
- Iida, S., Meyer, J. & Arber, W. *Plasmid* **1**, 357-367 (1978).
- Prentki, P., Karch, F., Iida, S. & Meyer, J. *Gene* **14**, 289-299 (1981).
- Ohtsubo, H. & Ohtsubo, E. *Proc. natn. Acad. Sci. U.S.A.* **75**, 615-619 (1978).

- Johnsrud, L. *Molec. gen. Genet.* **169**, 213-218 (1979).
- Fedoroff, N. V. *Cell* **16**, 551-563 (1979).
- Meyer, J., Iida, S. & Arber, W. *Molec. gen. Genet.* **178**, 471-473 (1980).
- Miller, J. H. *et al. J. molec. Biol.* **144**, 1-18 (1980).
- Shapiro, J. A. & MacHattie, L. A. *Cold Spring Harb. Symp. quant. Biol.* **43**, 1135-1142 (1979).
- Gottesman, M. M. & Rosner, J. L. *Proc. natn. Acad. Sci. U.S.A.* **72**, 5041-5045 (1975).
- De Bruijn, F. J. & Bukhari, A. I. *Gene* **3**, 315-331 (1978).
- Iida, S. & Arber, W. *Molec. gen. Genet.* **173**, 249-261 (1979).
- Grindley, N. D. F. & Sherratt, D. J. *Cold Spring Harb. Symp. quant. Biol.* **43**, 1257-1261 (1979).
- Harshey, R. M. & Bukhari, A. I. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1090-1094 (1981).
- Galas, D. J. & Chandler, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4858-4862 (1981).
- Shapiro, J. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1933-1937 (1979).
- Arthur, A. & Sherratt, D. *Molec. gen. Genet.* **175**, 267-274 (1979).
- Gill, R., Heffron, F., Dougan, G. & Falkow, S. *J. Bact.* **136**, 742-756 (1978).
- Heffron, F., McCarthy, B. J., Ohtsubo, H. & Ohtsubo, E. *Cell* **18**, 1153-1163 (1979).
- Bolivar, F. *Gene* **4**, 121-136 (1978).
- Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).

Promoters in the *E. coli* replication origin

Heinz Lother & Walter Messer

Max-Planck-Institut für molekulare Genetik, Abt. Trautner, Ihnestr. 63/73, D-1000 Berlin 33, FRG

Chromosomal replication in *Escherichia coli* initiates at a particular sequence in the parental DNA called the origin of replication (*oriC*) and proceeds bidirectionally¹. Initiation requires the interaction of various factors at a given time within the cell cycle, including RNA polymerase^{2,3}. Recently, minichromosomes have been constructed⁴⁻⁶ and partially sequenced^{7,8} whose replication proved also to be dependent on a rifampicin-sensitive host function in addition to the products of several initiation genes^{6,9}. Using one of these minichromosomes, pCM959^{6,8}, we have searched for promoters in and close to *oriC* to elucidate the role of RNA polymerase in the initiation of replication. We have found, by *in vitro* transcription experiments and RNA sequencing, that the DNA segment required for the start of bidirectional replication in *E. coli* contains two promoters arranged back-to-back. The location and arrangement of these promoters suggest an involvement in the initiation of replication.

Several complementary techniques were used to identify promoters likely to be involved in the initiation of replication. First we surveyed for promoters by mapping RNA polymerase binding sites present on pCM959 using electron microscopy. Three RNA polymerase binding sites, at nucleotide positions 267 ± 61, 791 ± 55 and 3,991 ± 68 of pCM959, mapped close to or within *oriC*; details will be published elsewhere¹⁰.

We used an *in vitro* transcription assay to determine the direction of transcription, the precise location of the transcription starts and the relative efficiencies of these promoters, which presumably have different functions. Various restriction fragments from pCM959 encompassing the three RNA polymerase binding sites of interest were used as templates for

purified RNA polymerase. The products of such reactions are shown in Fig. 1.

Figure 2 shows the alignment of transcripts that end at the termini of these restriction fragments (run-off transcripts). For example, using the *Hae*III fragment (position 3,970 to 636) as template, transcripts are obtained that are 220 and 325 nucleotides long (Figs 1b, 2). When this fragment is cleaved with *Hha*I (position 475) or *Hind*III (position 244), the 220 (215)-nucleotide transcript is still present, but instead of the 325-nucleotide transcript, a shorter length of 170 nucleotides is found, or, in the case of *Hind*III cleavage, the transcript is missing entirely (Figs 1a, 2). A 125-nucleotide transcript, however, is obtained if the *Hind*III-*Xho*I fragment (position 244 to 417) is assayed. The results obtained with the different fragments (Fig. 2) suggest that the RNA polymerase binding site observed at position 267 ± 61 is in fact due to two promoters situated back-to-back, which we designate *Pori-l* and *Pori-r*. According to this analysis the RNA origins should be located at position 170 ± 7 for *Pori-l* with counterclockwise transcription, at position 310 ± 10 for *Pori-r* with clockwise transcription, at position 765 ± 7 for the binding site at position 791 ± 55 with counterclockwise transcription, and at position 3,980 ± 10 for the binding site at 3,991 ± 68 with counterclockwise transcription.

To determine precisely the transcriptional starts, and to confirm the alignment, various 5'-end-³²P-₄-labelled transcripts obtained from restriction fragments encompassing both origin promoters were subjected to RNA sequence analysis¹¹⁻¹⁴. The RNA sequences were then compared with the DNA sequence. Transcripts originating from *Pori-l* started mainly with C at position 166 (Figs 3, 4). Transcripts originating from *Pori-r* started mainly with A at position 323 (Figs 3, 4); some transcripts from *Pori-r*, however, started with G at position 313 (complete sequencing data will be published elsewhere).

The directions of transcription determined for the promoters outside *oriC* agree with those inferred from the pCM959 nucleotide sequence (Fig. 3). These two promoters are followed by a Shine-Dalgarno sequence¹⁵ within the postulated distance¹⁶ of an open reading frame for proteins with molecular weights (MWs) of 16,000 (ref. 7) and 56,000, respectively (Fig. 3). The function of the 16,000-MW protein is unknown but it does not seem to have a major effect on minichromosome

Fig. 1 Electrophoretic analysis of transcripts synthesized by purified *E. coli* RNA polymerase holoenzyme from various restriction fragments encompassing *oriC*. The transcription experiments were carried out in a final volume of 200 μ l with 0.05–0.2 μ g purified DNA fragment and 5–15 μ g RNA polymerase (85 mol RNA polymerase per mol DNA fragment²⁸) in 40 mM Tris-HCl pH 8.0, 10 mM MgCl₂, 4 mM dithiothreitol, 400 μ M ATP and GTP, 100 μ M CTP and UTP + 50–100 μ Ci[α -³²P]UTP with *a*, 50 mM KCl and 100 μ g ml⁻¹ heparin^{29,30} or *b*, 50, 100 or 150 mM KCl (ref. 28). Reaction mixtures were preincubated for 10 min at 37 °C in the absence of *a*, UTP, CTP and heparin, or *b*, MgCl₂ (ref. 28). The complete reaction mixture was incubated for 20 min at 37 °C and the reaction stopped by addition of phenol. Samples were precipitated with ethanol, dissolved in 50% formamide, treated for 2 min at 100 °C, and subjected to electrophoresis in 8% (*a*) or 10% (*b*) polyacrylamide gels containing 7 M urea³¹. Numbers above the tracks specify the coordinates of the restriction fragments. Numbers alongside major transcripts represent their size in bases. ³²P-labelled DNA fragments served as length standards. Transcripts having no size assignments are those terminated at transcription termination signals and transcripts with aberrant electrophoretic behaviour, as evidenced by the nucleotide sequence of their 5'-terminal portions.

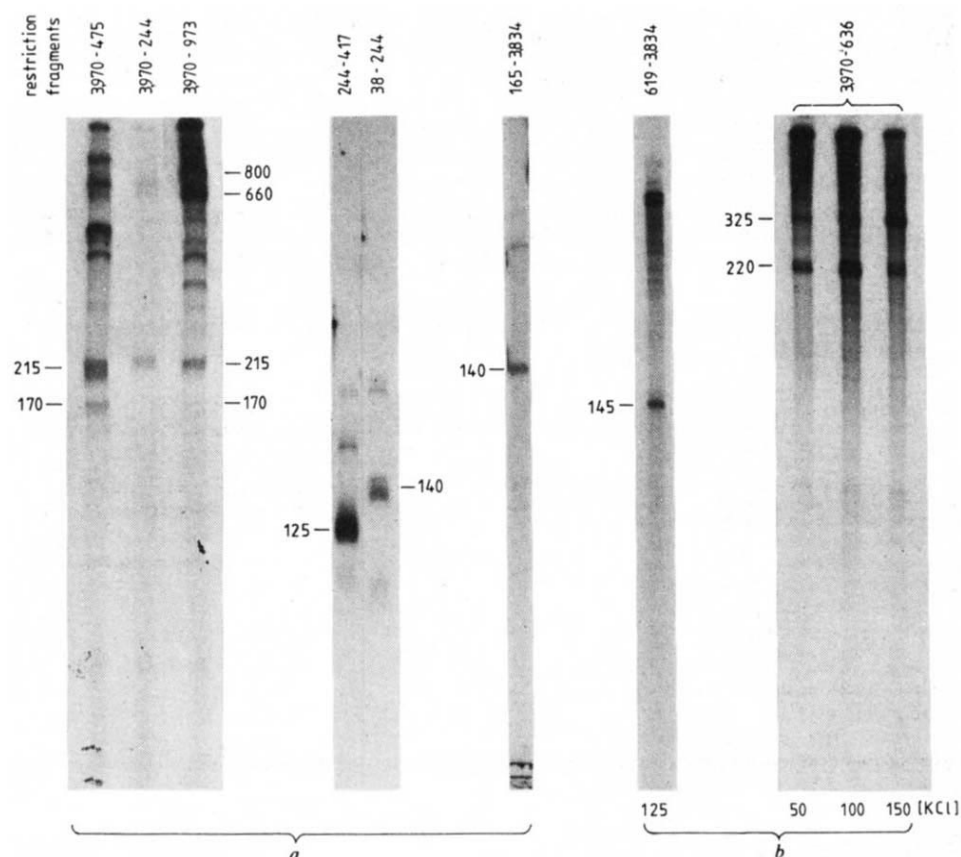
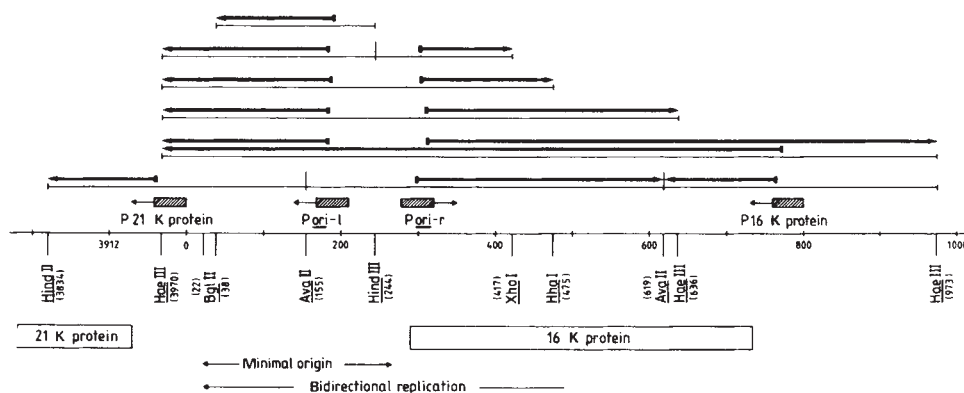


Fig. 2 Alignment of transcripts obtained from restriction fragments containing *oriC* and/or adjacent DNA segments. According to the length determination specified in Fig. 1, major transcripts (thick lines) were arranged by aligning their 3' ends with the ends of restriction fragments (thin lines). The positions of the promoters are indicated by cross-hatched boxes. Arrows indicate transcription directions. Numbers next to restriction sites within the physical map specify their coordinates. The open boxes beneath the physical map specify the coding regions of proteins. Both the limits of the minimal origin²³ and the region required for bidirectional replication²⁴ are shown. The *Hae*III (636) site is only present in a derivative of pCM959.



replication¹⁷, even though its coding region overlaps partially with that of transcripts from *Pori-r*. The 56,000-MW protein is a fusion product of a 21,000- and a 35,000-MW partial peptide as its coding region traverses the point of circularization of pCM959 (data not shown). The 21,000-MW partial peptide may be the beginning of a 70,000-MW protein designated *gid*, which was mapped in this region¹⁷. Both a 16,000- and a 70,000-MW protein have been seen after infection of minicells carrying the λ repressor-overproducing plasmid pGY101¹⁸ with λ vectors into which different minichromosomes were inserted (unpublished results), and in cells carrying minichromosomes¹⁹. Both these promoters were also demonstrated *in vivo*¹⁹.

Although these observations suggest a linkage between transcription and translation for the two promoters outside *oriC*, a

coupling of these two processes for the two *oriC* promoters seems unlikely. Nonsense codons are present in all reading frames, and the longest possible peptides which can be deduced from the nucleotide sequence downstream of *Pori-l* and *Pori-r* are 27 amino acids long (130–50) and 55 amino acids long (351–525), respectively (Fig. 3). Such peptides have so far not been identified. The origin promoters seem to be as active *in vitro* as the 16,000-MW protein promoter and far more active than the 21,000-MW partial protein promoter, which expresses transcripts that are barely above background. Therefore we suggest that these promoters are also used *in vivo*, but for a purpose other than protein biosynthesis.

When the nucleotide sequences of the two origin promoters are compared, we observe extended sequence homologies

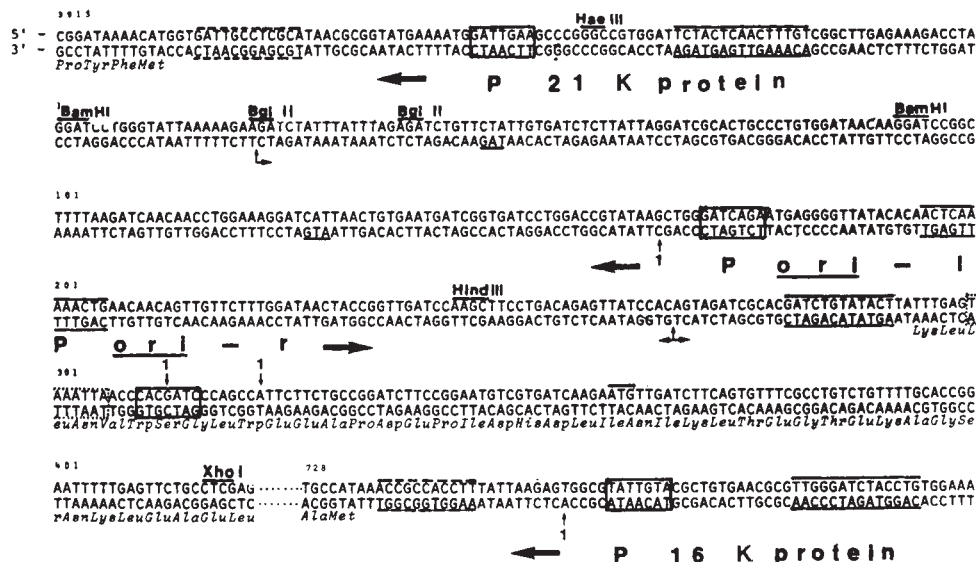
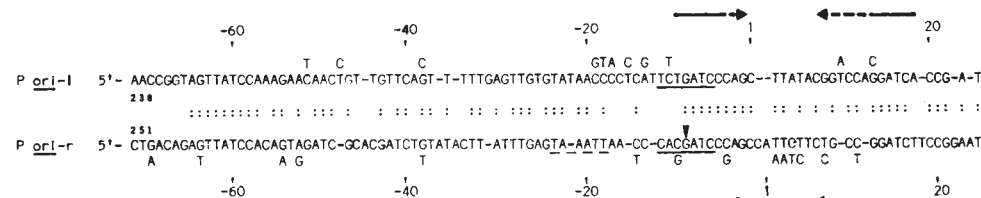


Fig. 4 Comparison of origin promoters of *E. coli* and *S. typhimurium*. The sequences of the noncoding strands starting at base 238 and 251, respectively, are compared. Numbers above and below the sequences refer to nucleotide positions relative to the transcriptional start. The 'secondary' Pribnow box occurring in *Pori-r* is marked by dashes; ▼ represents its corresponding 5' start nucleotide. The corresponding *S. typhimurium* sequence²¹ is shown above or below the *E. coli* sequence only where there are sequence differences. : indicate identical bases in the two *E. coli* promoter sequences. Note that most changes within the promoter sequences are found in regions which are also more variable in other promoters. The changes are located between the Pribnow boxes (underlined regions) and around base -20 (ref. 20). Other changes occur within the loop of a putative operator structure indicated by inverted arrows.

around position -50 to -70 and close to the transcription starts (Fig. 4). The region preceding the actual promoter sequence is known to be important for positive control in the *lac*, *gal* and other systems²⁰. This sequence homology is also evident when these regions are compared with the corresponding segment in the *Salmonella typhimurium* origin (ref. 21; Fig. 4) and with the other origins of Gram-negative bacteria so far sequenced, for example *Klebsiella pneumoniae*, *Enterobacter aerogenes* and *Erwinia carotovora*²². Thus an attractive hypothesis is that such RNA synthesis occurs in all these Gram-negative bacteria and that it is subject to positive control in all cases. The second conserved sequence homologies are close to the transcription starts. These regions of homology can form imperfect palindromes and may therefore represent sites of interactions, occurring on both promoters, with a hypothetical repressor.

The minimal DNA segment required for origin function extends from position 22 to position 267 (ref. 23; Fig. 3) whereas *oriC*, the origin required for bidirectional replication, has been shown to extend beyond position 417 (ref. 24). *Pori-l* is thus completely contained in the DNA segment that is sufficient for leftward unidirectional replication. *Pori-r* is located in that part of *oriC* which is, in addition, required for bidirectional replication. This arrangement suggests that the origin promoters are responsible for the synthesis of RNA primers required for the inception of the leading DNA strands. However, we cannot rule out the possibility that *Pori-l* and *Pori-r* promote, together or exclusively, transcriptional activation²⁵ or that one or both transcripts have a regulatory role similar to that of the short RNA molecules found in some plasmid systems^{26,27}.

Fig. 3 Location of the four promoters identified *in vitro* within and close to *oriC*. Both the DNA sequence and its corresponding amino acid sequence are shown. Note that nucleotide 313 serves also as a transcript start. Its corresponding Pribnow box²⁰ is marked by a dotted line. ↑ Transcription start; boxes indicate Pribnow boxes as inferred from the transcription starts²⁰; lines above and below the sequences represent -35 regions²⁰; broken lines indicate Shine-Delgarno sequences¹⁵. →, Direction of transcription; ↔, limits of the minimal origin segment²³. Start and stop codons downstream of *Pori-l* and *Pori-r* are underlined.



Received 16 July; accepted 6 October 1981.

1. Masters, M. & Broda, P. *Nature new Biol.* **232**, 137-140 (1971).
2. Lark, K. G. *J. molec. Biol.* **64**, 47-60 (1972).
3. Messer, W. *J. Bact.* **112**, 7-12 (1972).
4. Yasuda, S. & Hirota, Y. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5458-5462 (1977).
5. Messer, W. *et al. Molec. gen. Genet.* **162**, 269-275 (1978).
6. Meyenburg, K. V. *et al. Cold Spring Harb. Symp. quant. Biol.* **43**, 121-128 (1979).
7. Sugimoto, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 575-579 (1979).
8. Meijer, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 580-584 (1979).
9. Lother, H. *et al. ICN-UCLA Symp. molec. cell. Biol.* **21** (in the press).
10. Morelli, G. *et al. Molec. gen. Genet.* (in the press).
11. Donis-Keller, H., Maxam, A. M. & Gilbert, W. *Nucleic Acids Res.* **4**, 2527-2538 (1977).
12. Simonsoit, A., Brownlee, G. G., Brown, R. S., Rubin, J. R. & Guillely, H. G. *Nature* **269**, 833-836 (1977).
13. Lockard, R. E. *et al. Nucleic Acids Res.* **5**, 37-56 (1978).
14. Krupp, G. & Gross, H. J. *Nucleic Acids Res.* **6**, 3481-3490 (1979).
15. Shine, J. & Dalgarno, L. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1342-1346 (1974).
16. Steitz, J. A. & Jakes, K. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4734-4738 (1975).
17. Meyenburg, K. v. & Hansen, F. G. *ICN-UCLA Symp. Molec. Cell. Biol.* **19**, 137-159 (1980).
18. Reeve, J. N. *Nature* **276**, 728-729 (1978).
19. Hansen, F. G., Koefoed, S., Meyenburg, K. v. & Atlung, T. *ICN-UCLA Symp. Molec. Cell. Biol.* **21** (in the press).
20. Rosenberg, M. & Court, D. A. *Rev. Genet.* **13**, 319-353 (1979).
21. Zyskind, J. W. & Smith, D. W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2460-2464 (1980).
22. Zyskind, J. W., Harding, N. E., Takeda, Y., Cleary, J. M. & Smith, D. W. *ICN-UCLA Symp. Molec. Cell. Biol.* **21** (in the press).
23. Oka, A., Sugimoto, K., Takanami, M. & Hirota, Y. *Molec. Gen. Genet.* **178**, 9-20 (1980).
24. Meijer, M. & Messer, W. *J. Bact.* **143**, 1049-1053 (1980).
25. Dove, W. F., Inokuchi, H. & Stevens, F. in *The Bacteriophage Lambda* (ed. Hershey, A. D.) 747-771 (Cold Spring Harbor Laboratory, New York, 1971).
26. Itoh, T. & Tomizawa, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2450-2454 (1980).
27. Rosen, J., Ryder, T., Ohtsubo, H. & Ohtsubo, E. *Nature* **290**, 794-797 (1981).
28. Otsuka, A. & Abelson, J. *Nature* **276**, 689-694 (1978).
29. Meyer, B. J., Kleid, D. G. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4785-4789 (1975).
30. Zillig, W. *et al. Cold Spring Harb. Symp. quant. Biol.* **35**, 47-58 (1970).
31. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).

Molecular dynamics of hydrogen bonds in bovine pancreatic trypsin inhibitor protein

Michael Levitt

Department of Chemical Physics, Weizmann Institute, Rehovot, Israel

Molecular dynamics simulations starting from the X-ray structure¹ of bovine pancreatic trypsin inhibitor protein (BPTI) reveal that the different hydrogen bonds observed in the X-ray coordinates have different stabilities as indicated by the mean lengths and length fluctuations. The most stable hydrogen bonds involve the hydrogen atoms observed to exchange most slowly with solvent in NMR experiments², and to have the shortest lengths in the X-ray structure. This agreement between calculation and experiment suggests that molecular dynamics simulations can complement X-ray studies by providing reliable information about the rates and pathways of conformational changes about the mean positions observed in protein crystals.

The results of dynamics simulations depend on the molecular potential energy from which the forces acting on the atoms are calculated. As all force calculations on proteins³⁻⁵ use an approximate empirical potential energy function and omit the thousands of surrounding solvent molecules, it is essential to check the results quantitatively before using such theoretical

methods to calculate quantities that cannot be measured. Two observable properties of globular proteins seem particularly well suited to such a test: (1) the amplitudes of atomic vibration available from the highest-resolution X-ray studies can be compared with the amplitudes computed from a dynamics run; and (2) the rates at which hydrogen atoms in the protein exchange with deuterium atoms in solution, as measured by ¹H NMR, can be compared with the computed fluctuations in the hydrogen bond lengths. Northrup *et al.*⁶ have shown that the observed and calculated atomic vibration amplitudes yield similar pictures for the atomic mobility in the protein ferrocytochrome *c*.

Here I analyse the results of molecular dynamics simulations on the protein BPTI in terms of the length fluctuations of hydrogen bonds. BPTI is the obvious choice for this study as the rates of exchange of individual protons are not available for other globular proteins². In the calculation, the equations of motion in the protein force field used previously^{5,7} are solved by the Beeman⁸ method of numerical integration with a time step of 2×10^{-15} s. Rapid equilibration is achieved by starting from a completely relaxed structure generated by extensive conjugate gradient minimization⁵ (up to 3,000 energy evaluations) from the X-ray structure and then gradually increasing kinetic energy by small random impulses to give an equilibrium velocity distribution (gaussian) with no local 'hot spots'.

The mean lengths and length fluctuations of hydrogen bonds calculated from a 56-ps molecular dynamics simulation (see Table 1) are insensitive to the part of the trajectory from which they are calculated (correlation coefficient of mean lengths in MD1 and MD2 is 0.98), indicating that the trajectory is well equilibrated for calculation of these quantities. The different hydrogen bonds in the structure show systematic differences in mean lengths and length fluctuations allowing all 24 hydrogen bonds in Table 1 to be classified as 'stable', if the mean length (the O—H distance) is < 1.90 Å and the r.m.s. length fluctuation

Table 1 Calculated and observed properties of hydrogen bonds

Secondary structure and hydrogen bond	Mean length (Å)		Mean length (Å)		Class	X-ray length (Å)	log ₁₀ ¹ H exchange rate (min ⁻¹)
	MD1 (1-24 ps) Mean	Amp.	MD2 (32-56 ps) Mean	Amp.			
α-helix							
Pro 2, O—Cys 6, H	2.11	0.32	2.23	0.47	U	2.25	0
Asp 3, O—Leu 7, H	2.15	0.33	2.12	0.56	U	2.17	0
β-hairpin							
Thr 11, O—Gly 36, H	1.83	0.11	1.81	0.10	S	1.97	-2
Gly 36, O—Ala 16, H	4.01	0.50	4.10	0.37	U	2.19	0
Tyr 35, O—Ile 18, H	2.64	0.63	2.38	0.65	U	2.21	-5
Ile 18, O—Tyr 35, H	1.79	0.10	1.78	0.09	S	1.68	-5
Phe 33, O—Arg 20, H	1.80	0.09	1.80	0.10	S	2.05	-7
Arg 20, O—Phe 33, H	1.79	0.09	1.80	0.10	S	2.00	-6
Phe 45, O—Tyr 21, H	1.81	0.10	1.82	0.10	S	1.71	-6
Tyr 21, O—Phe 45, H	1.77	0.08	1.79	0.09	S	1.76	-6
Gln 31, O—Phe 22, H	1.82	0.10	1.82	0.11	S	1.90	-6
Phe 22, O—Gln 31, H	1.76	0.08	1.77	0.08	S	1.96	-5
Leu 29, O—Asn 24, H	1.81	0.10	1.85	0.15	S	2.02	1.81
Asn 24, O—Gly 28, H	1.82	0.10	1.82	0.10	S	2.06	0
α-helix							
Ser 47, O—Cys 51, H	2.74	0.60	2.33	0.45	U	2.13	0
Ala 48, O—Met 52, H	1.89	0.21	1.88	0.18	S	1.82	-4
Asp 49, O—Arg 53, H	2.73	0.62	2.56	0.64	U	2.12	0
Asp 50, O—Thr 54, H	3.28	0.57	3.07	0.44	U	2.13	0
Cys 51, O—Cys 55, H	1.81	0.12	1.82	0.16	S	1.83	-3
Met 52, O—Gly 56, H	3.35	0.44	3.41	0.32	U	1.94	0
Side chain/main chain							
Glu 7, O—Asn 43, HD1	1.93	0.32	1.82	0.13	S	1.84	-2
Tyr 23, O—Asn 43, HD2	1.83	0.13	1.86	0.14	S	2.03	-5
Asn 24, OD1—Lys 26, H	1.97	0.20	1.99	0.22	U	2.65	0
Asn 43, OD1—Tyr 23, H	1.81	0.11	1.80	0.10	S	1.82	-7

The amplitude (amp.) is the r.m.s. deviation of the hydrogen bond length from the mean value for the specified part of the trajectory. log₁₀ ¹H exchange rates are taken from Table 1 in ref. 2. The rapidly exchanging protons are not seen in the experiment and are given the rate of 1 min⁻¹, corresponding to that of an exposed amide group in model peptides¹⁴. U, unstable; S, stable.

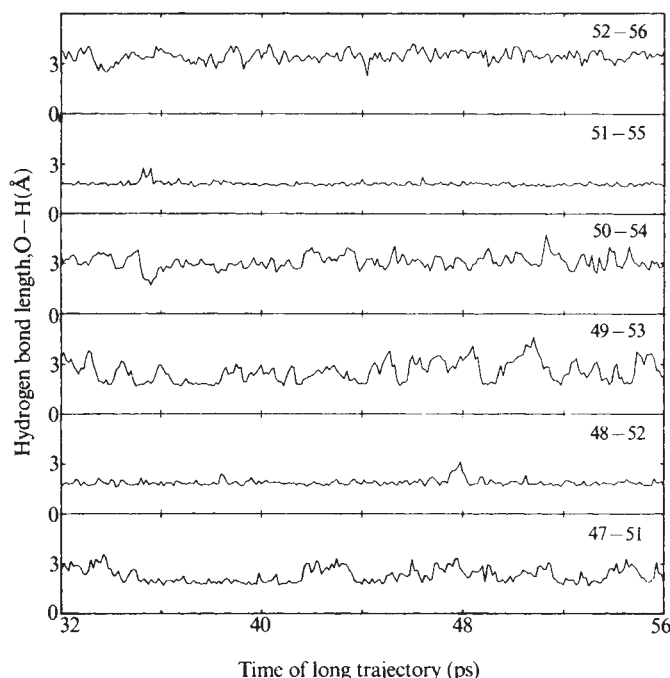


Fig. 1 The time variation of the six O—H distances of the peptide hydrogen bonds observed in the C-terminal α -helix of BPTI during the period of 32–56 ps of the trajectory. A well-formed hydrogen bond has a length of 1.7–2.2 Å.

(amp.) <0.22 Å (marked 'S' in Table 1), or 'unstable' otherwise (marked 'U'). Comparison with the experimentally determined X-ray lengths¹ shows that the 'stable' hydrogen bonds are significantly shorter: 15 of the 16 X-ray lengths <2.07 Å correspond to hydrogen bonds defined as stable by the independent criteria given above. Comparison with the ¹H proton NMR exchange rates measured for peptide protons² shows that the 'stable' hydrogen bonds have significantly slower proton exchange rates: 14 of the 15 most slowly exchanging protons found experimentally correspond to stable hydrogen bonds (S in Table 1).

This agreement between the classification of the 24 hydrogen bonds on the basis of the dynamics simulation and either the X-ray lengths or the NMR exchange rates is highly (99.99%) significant. If the 24 hydrogen bonds were to be randomly assigned as 'S' or 'U' subject to the requirement that there be exactly 15 'S' assignments, there would be $24!/(15!9!)$ distinct arrangements. One of these would match the NMR exchange rates exactly and 9×15 would have two mismatches, as is found here. The probability of getting chance agreement as good as that seen here is $(9 \times 15 + 1)/(24!/(15!9!)) = 136/1,307,504 = 0.0001$.

Despite this good agreement, the correlation between the mean lengths (or length fluctuations) and the actual exchange rates of individual hydrogen bonds is poor (for example, both Phe 33, O—Arg 33, H and Thr 11, O—Gly 36, H are correctly classified as stable but have exchange rates that differ by five orders of magnitude). Thus although the calculation can successfully predict which of the hydrogens will exchange slowly, it cannot predict the actual exchange rates. This is to be expected as hydrogen exchange occurs on the time scale of minutes and depends on many factors besides the stability of the hydrogen bond, including local unfolding, solvent penetration and catalysis by neighbouring groups.

Why are certain hydrogen bonds unstable as measured by the three independent criteria of dynamics simulation, X-ray lengths and NMR exchange rate? One explanation is that the

potential energy of certain hydrogen bonds is high (less favourable) due to local conformational strain. Another is that certain hydrogen bonds fluctuate more to increase their conformational entropy and so reduce the free energy of the system. The entropy gained from larger fluctuations depends on the density of packing around the hydrogen bond in question: if there is no space to move into, there is nothing to be gained from large fluctuations. Support for this idea comes from the accessibilities of the bonds (calculated as the sum of the static accessible surface areas⁹ of the O and N atoms in the X-ray structure): 5 of the 9 'U' hydrogen bonds in Table 1 are exposed (3, O—7, H = 8.3 Å²; 36, O—16, H = 3 Å²; 24, O—28, H = 0.3 Å²; 49, O—53, H = 7.2 Å²; and 50, O—54, H = 37.8 Å²) whereas only 2 of the 15 'S' hydrogen bonds are exposed (48, O—52, H = 7.7 Å² and 24, OD1—26, H = 9.8 Å²).

Figure 1 shows the time variation of the six peptide hydrogen bond lengths in the C-terminal α -helix. The two stable hydrogen bonds (Ala 48, O—Met 52, H and Cys 51, O—Cys 55, H) vibrate mostly about the length value of 1.8 Å, but occasionally show increases in length of up to 3 Å for periods of ~ 0.5 ps; these fluctuations do not involve the correlated breaking of hydrogen bonds adjacent along the α -helix. The four unstable hydrogen bonds (Ser 47, O—Cys 51, H; Asp 49, O—Arg 53, H; Asp 50, O—Thr 54, H; and Met 52, O—Gly 56, H) show larger length fluctuations, but a good hydrogen bond is sometimes formed for a few picoseconds. The same type of dynamic behaviour is seen for the other hydrogen bonds and for other periods of the trajectory.

Discussion of the broad range of rates with which protons exchange in globular protein has in the past focused on how the particular groups contact the solvent^{10–12,13}. My study emphasizes another factor showing that the intrinsic stability of different hydrogen bonds differs and that the stable hydrogen bonds exchange more slowly.

This work differs from previous studies of BPTI dynamics^{3,4} in two important respects. (1) I allow hydrogen bonds to be formed between all oxygen and hydrogen atoms that are close together in space. As the previous studies used a fixed list of hydrogen bonds taken from the X-ray data¹, questions concerning hydrogen bond fluctuations could not be answered. (2) I use a different potential energy function, initially calibrated on the observed unit cell parameters of hydrocarbon, amide and amino acid crystals and then tested for its ability to give an equilibrium conformation close to the observed X-ray structure⁵. The present simulations do, however, have a serious shortcoming in the omission of the thousands of water molecules surrounding the protein molecule: 2,700 water molecules would have to be included to give a 0.02 M solution (very concentrated by experimental standards), increasing the computer time for one time step by a factor of ~ 10 . This more realistic calculation is now in progress.

This work was partially supported by NSF award PCM-7808029. I thank Professor F. H. C. Crick for encouragement and critical discussion, and Dr R. Feldmann for making a film of the trajectory.

Received 12 February; accepted 22 September 1981.

1. Deisenhofer, J. & Steigemann, W. *Acta crystallogr.* **B31**, 238–250 (1974).
2. Wütrich, K. & Wagner, G. *J. molec. Biol.* **130**, 1–18 (1979).
3. McCammon, J. A., Gelin, B. R. & Karplus, M. *Nature* **267**, 585–589 (1977).
4. McCammon, J. A., Wolynes, P. G. & Karplus, M. *Biochemistry* **18**, 927–942 (1979).
5. Levitt, M. in *Protein Folding* (ed. Jaenicke, R.) 17–39 (Elsevier, Amsterdam, 1980).
6. Northrup, S. H. *et al. Nature* **287**, 659–660 (1980).
7. Levitt, M. *J. molec. Biol.* **145**, 251–263 (1981).
8. Beeman, D. *J. comp. Phys.* **20**, 130–139 (1976).
9. Lee, B. & Richards, F. M. *J. molec. Biol.* **55**, 379–400 (1971).
10. Englander, S. W. & Englander, J. J. *Meth. Enzym.* **49**, 24–39 (1978).
11. Wütrich, K. *et al. Biophys. J.* **10**, 549–560 (1980).
12. Woodward, C. K. & Hilton, B. D. *Biophys. J.* **10**, 561–571 (1980).
13. Richards, F. M. *et al. in Protein Folding* (ed. Jaenicke, R.) 117–129 (Elsevier, Amsterdam, 1980).
14. Englander, S. W., Downer, N. W. & Teitelbaum, H. A. *Rev. Biochem.* **41**, 903–924 (1972).

BOOK REVIEWS

Environmental impact re-assessment?

David Pearce

THE current fashion for environmental impact assessment (EIA) is partly explained by the continuing force of the environmental protection movement in Western countries. That movement is now under severe pressure from economic recession, and there are signs that impact assessments themselves will play a decreasing role in planning and development. Certainly, this is the message that emerges from the USA, where the emphasis is switching back to the costs of environmental protection. This reverse swing, no sign of which is even hinted at in the two volumes under review, is both unfortunate and misguided. But the environmental movement itself bears some of the responsibility. In their indecent rush to find an approach (some might say any approach) to the integration of environmental factors with the social and economic dimensions of development, environmentalists have produced a jumble of methodologies all too few of which have any rational grounding.

In essence, an EIA should identify all environmental and social impacts from a development. It may then leave them qualitatively recorded, or go further and measure them in physical terms or as "scores" reflecting a judgement of whether that impact is good, bad, moderate or serious. Neither *Project Appraisal and Policy Review*, which is a collection of essays on the state of play in EIA, nor *Environmental Impact Assessment*, the proceedings of the 1979 meeting of the UN Economic Commission for Europe (ECE), will aid anyone unfamiliar with the methodologies and techniques used for EIA. The ECE volume does, however, highlight the prevailing chaos whereby almost anything that quantifies, indicates or even mentions "impacts" is classified as an EIA. The conference lamented this state of affairs and one of its recommendations is for systematization. But there is a surprising lack of self-criticism in both books. Some of the scoring methods, for example, actually offend logic by placing cardinal numbers on ordinal rankings. Moreover, the call for systematization is negated by fairly regular references to the "flexibility" of EIA, as if this is a virtue rather than a euphemism for arbitrary procedures.

Both volumes would have greater claim to authority had they included contributions from economists. One example will suffice. In the USA the National Environmental Protection Act mandated EIA for public sector investments. A fairly

Project Appraisal and Policy Review. Edited by Timothy O'Riordan and W. R. Derrick Sewell. Pp. 304. ISBN 0-471-27853-X. (Wiley: 1981.) \$36, £13. *Environmental Impact Assessment*. Proceedings of a Seminar of the United Nations Economic Commission for Europe. Pp.368. ISBN 0-08-024445-9. (Pergamon: 1981.) \$55, £23.

traditional cost-benefit analysis was to be included. Without it, one may observe, it is not clear how anyone is to decide whether the investment is worth while at all. Yet O'Riordan and Sewell speak of the over-narrow concern in cost-benefit analysis with "economic efficiency" while their co-authors speak of the deficiencies of planning procedures that cannot debate national "need", which is what economic efficiency is all about.

A further flaw is that the value system underlying EIA is not investigated in either book. Cost-benefit analysis attempts to reflect an aggregate of individuals' preferences. In EIA, though it is hardly acknowledged, it is the *planners'* preferences that count. Possibly the unease which a democrat should feel in this respect explains the emphasis given by some of the authors in the O'Riordan and Sewell volume to the need to link EIA to "public involvement", although one cannot be quite sure who the "public" are. An EIA that finally emerges as a matrix of impacts does at least provide information for local and central government, as the authors correctly emphasize. But that same matrix — consisting of incommensurable impacts, subjective assessments by planners and consultants, objective physical measures

which may or may not be disputed by experts — cannot then answer questions of social worth unless reduced to commensurable terms with the demands the development makes on resources, i.e. its money cost. Moreover, EIA is of limited value in assisting choice between *alternatives* unless the impacts in the matrix for one assessment are systematically greater or less than the impacts in the assessment of the alternative.

The most provocative essay in the O'Riordan and Sewell volume, by Nick Abel and Michael Stocking, questions whether EIA has any relevance to less developed countries. Their argument is basically that EIA enshrines Western values, thus providing inappropriate evaluations of development projects, for example by implicitly equating low economic output per capita with "backwardness". Their strictures are familiar but worthy of careful examination. Anyone who has worked in low per capita income countries will doubt if Westernized assessment procedures are quite so irrelevant, since the general aim of income expansion is wholly warranted. But there is an unquestionable need to place EIA and other appraisal techniques in the political context of the country in question.

To anyone looking for an "update" on the state of play in EIA these are useful volumes, *Project Appraisal and Policy Review* especially so. Sadly, however, neither book has come to terms with the underlying questions which any project appraisal needs to ask. □

David Pearce is Professor of Political Economy at the University of Aberdeen.

Lives and work of the nervous doctors

W.F. Bynum

The Doctrine of the Nerves: Chapters in the History of Neurology. By John D. Spillane. Pp. 467. ISBN 0-19-261135-6. (Oxford University Press: 1981.) £25, \$50.

THE title of Dr Spillane's volume comes from Thomas Willis's (1621–1675) original definition of "neurologie" as the branch of medicine dealing with "the doctrine of the nerves". The word, of course, has since changed in meaning, and long before Willis the central nervous system and its associated nerves had been recognized as a

crucial custodian of our uniqueness: the receptacle of our sensations, originator of our movements, storehouse of our memories, the seat of our souls. Consequently, the study of the structure, functions and derangements of the nervous system has a rich history, as this elegantly written and beautifully illustrated book attests.

Unlike Caesar's Gaul, Dr Spillane's book is divided into four parts. He describes the classical foundations (Galen, Vesalius, Willis), the eighteenth-century

obsession with the nervous system as the origin of most disease, the nineteenth-century establishment of experimental neuroscience and, finally, the flowering of clinical neurology in the late nineteenth century. The approach is primarily biographical and expository, which permits selectivity and places the value of the work on its individual sections rather than on any unifying synthesis. As the sub-title suggests, the book is a series of chapters rather than an integrated history.

This approach has its limitations and its strengths. Among the latter is the way it permits Spillane to bring his own neurological knowledge to bear on the ideas of his intellectual forebears. This is particularly true for the more recent neurologists such as G.B.A. Duchenne, W.R. Gowers and Silas Weir Mitchell, men whose names are invoked more often than their works are read. Spillane has also rehabilitated several men whose contributions are often forgotten: John Cooke (1756–1838) and John Russell Reynolds (1828–1896) fall into this category. Nearly 200 illustrations are neatly integrated into the narrative, and production standards are high.

On the other hand, such an approach can be self-indulgent and historically impoverishing. There is little social context for the individuals discussed, little consideration of the wider professional concerns involved in the emergence of neurology as a speciality, and little attempt to integrate Spillane's principal figures into mainstream medical history. Furthermore, the focus is too rigidly Anglo-American, with occasional excursions into France but little on the equally important German scene. M.H. Romberg gets some consideration, probably because his book was translated into English, but the earlier work of men such as K.F. Burdach, and the rich neurological and neuropsychiatric tradition created by Monakow, Westphal, Korsakoff, Benedikt, Wernicke and others are essentially neglected. It is also historically distracting to concentrate on organic diseases of the nervous system at the expense of "functional" disorders, hysteria for example, which so preoccupied physicians such as Weir Mitchell and Russell Reynolds. Aphasia and the neurological consequences of tertiary syphilis also get curiously brief attention.

It may seem ungenerous to carp at what Dr Spillane has failed to give us, when he has given us so much. For his book is a pleasure to read, and the copious use of quotation gives it the flavour of a source-book in the history of his discipline. It distils the secondary literature on many pioneers in neurology and the neurosciences. But it is decidedly chapters in, rather than a history of, neurology. □

W.F. Bynum is at the Wellcome Institute for the History of Medicine, London, and a co-editor of the Dictionary of the History of Science (Macmillan/Princeton University Press, 1981).

Ecology: much to be done in the North

Stanwyn G. Shetler

The Boreal Ecosystem. By James A. Larsen. Pp.500. ISBN 0-12-436880-8. (Academic: 1981.) \$45, £29.80.

ECOLOGICAL studies of the boreal forest region have proliferated in the past few decades, and synthesis of them is certainly overdue. Larsen's impressive and scholarly treatise is not all we might have hoped for, but it will be welcomed eagerly by all boreal biologists and allied students of the North.

The book proceeds from a consideration of the glacial–postglacial history of boreal vegetation to discuss soils, climate, plant communities, various processes, and the economic utilization and management of the boreal forest ecosystem. Although a substantial portion is devoted to presenting some of the author's own data from his 20 summers of fieldwork, overall the book is a review and analysis of other published works.

Despite the ambitious title, the work is less than a definitive treatise on the boreal forest biome. This is not a synthesis for animal ecologists — except for one interesting but general chapter on animal populations ("The Trophic Pyramid"), *The Boreal Ecosystem* deals with plants. Geographically, the author confines his treatment largely to North America, with the Canadian sector, where he has had the most experience, receiving the most detailed consideration. Eurasia, with two-thirds of the world's boreal forest area, is compared briefly here and there, and the extensive Scandinavian and Russian boreal ecological literature is only lightly covered. The perfunctory paragraph on the Appalachian extension of the boreal forest, which long has intrigued American ecologists, is disappointing.

The study of boreal ecology, as ecology generally, often has been hindered by differences in method and terminology. Fortunately, Larsen wastes no time in equating "boreal forest" in the American sense with "taiga" in the Russian sense; thus he does not perpetuate the erroneous notion commonly encountered among North American botanists that taiga is forest-tundra. Probably no two ecologists use the term "muskeg" in precisely the same sense, and such pivotal terms as this, which the author uses throughout, should have been more explicitly defined at the outset.

Larsen, a modern quantitative plant ecologist, begins and ends with an expression of faith in modelling and systems analysis, and he never misses a chance to demonstrate the superiority of the continuum concept and method in analysing boreal forest communities. Continua do not reduce easily to definable systems, however. If he does anything with his wealth of data and description, he convinces us that there are *many* boreal forests

and environments — in short, many ecosystems, which form a baffling continuum from the tundra complexes in the north to the temperate complexes in the south. Though he drops generalizations here and there throughout the book, most never quite jell, and, while remaining optimistic about future breakthroughs, he is the first to concede that we have not begun to know how to elaborate a workable model or to put numbers into the equations.

We cannot hold him accountable for the untidy nature of ecological science, and we can applaud his quest for the grand synthesis. But Larsen never quite rises above description, in spite of his ecosystem point of view and emphasis on dynamics and process. Nonetheless, he does bring many fresh insights and promising quantitative methods to bear on some of the most fascinating but perplexing and intractable ecological riddles of the North: vegetation history, northern forest limit, nature of succession, permafrost, broad-scale vegetational uniformity masking great local variation, to name a few. His analyses of the forest boundary communities and environmental conditions (particularly the frontal air mass dynamics) and of the inadequacies of classic successional concepts are especially useful and enlightening.

Much of the book is too detailed and heavy-going for quick and easy comprehension, and Larsen's treatise comes off more as a compilation and review than as a challenging new synthesis. Nevertheless, as a compendium on the state of knowledge of the North American boreal forest and its ecotones to the north and south, this unique book is the best and most complete single volume available. □

Stanwyn G. Shetler is a Curator of the Department of Botany at the Smithsonian Institution, Washington DC.

Evolution writ small

R.P. Ambler

Biochemical Evolution. Edited by H. Gutfreund. Pp.368. ISBN hbk 0-521-23549-9; ISBN pbk 0-521-28025-7. (Cambridge University Press: 1981.) Hbk £30, \$69.95; pbk £12.50, \$24.95.

THERE are two ways in which evolution can be studied at the biochemical level. One is by the traditional comparative approach, epitomized 40 years ago by Baldwin's *An Introduction to Comparative Biochemistry* (Cambridge University Press, 1940). The other is through molecular

genetics, following changes in gene sequence, number and position, while speculating about the accompanying effects on gene products, and studying functional change through enzymology or population genetics.

These approaches are illustrated by the nine essays that make up the present volume. Pink's excellent chapters on the major histocompatibility antigens and on immunoglobulins illustrate the power of modern molecular genetics in systems where comparative information is very limited. Pink is well aware of the need to consider function in evolving macromolecules, although in the immunoglobulins relevant information is limited. Is it even necessary (p.249) to accept that V gene proliferation is under selective control? In contrast, Crecitelli's account of the vertebrate visual pigments has very little to do with evolution. Weeds and Wagner's chapter on muscle contraction considers Kretsinger's and Collins' speculations about the universality of the structure of the calcium-binding site in contractile and other proteins, but only touches on the question of the multiplicity of actin and myosin genes.

Rao, Hall and Cammack illustrate some of the weaknesses of the comparative approach in their chapter on the photosynthetic apparatus, in which the now conventional attempt is made to derive from a synthesis of the properties of present-day organisms a synoptic account of the history of metabolism on Earth. In contrast, Clarke, in an essay that has a wider scope than its title "Enzymes in Bacterial Populations" suggests, is wise and sceptical, for example in pointing out that early metabolic pathways, like organisms, may have become extinct. In her final sentence she emphasizes the speed with which microbial evolution is observed to occur, a paradox when contrasted with speculation that particular present-day bacteria are recognizably similar to pre-Cambrian organisms.

Peacocke contributes a useful chapter on methods of data handling for phylogenetic trees. The longest section in the book is from Schuster, who considers various aspects of prebiotic evolution. After a long explanation of hypercycles, he concludes that

many plausible models (for the origin of the genetic code and the formation of protocells) can be proposed, but it is very hard to decide between them or even to suggest experiments to prove or disprove them.

Overall, the book suffers from insufficient editorial guidance on chapter content and emphasis, and the introductory chapter by Gutfreund does little to link the rest of the book together or to point the way of future experiment or speculation. □

R. P. Ambler is a Reader in Molecular Biology at Edinburgh University.

All angles on contemporary cosmology

Bernard J.T. Jones

The Isotropic Universe. Monographs on Astronomical Subjects, 7. By D.J. Raine. Pp.253. ISBN 0-85274-370-X. (Adam Hilger, Bristol/Heyden, Philadelphia: 1981.) £19.50, \$49.

EDWIN Hubble made a number of discoveries of fundamental importance in establishing our picture of the Universe, foremost amongst which was the realization that the Universe is expanding. On the simplest level of interpretation this astonishing fact demanded an origin for the Universe at a finite time in our past — a "big bang". Of almost equal importance was the realization that the Universe looks the same at great distances in whatever direction one observes: this homogeneity and isotropy fitted in well with the simple cosmological models deduced by Friedman and Lemaître from Einstein's general theory of relativity. These observations led to one of the great controversies of modern science: the issue was whether the Universe really had a singular origin (the big bang theory), or whether matter was being continuously created so as to make up the loss due to the expansion (the steady-state theory). In 1965 the debate was resolved with the discovery of the cosmic microwave background radiation by Penzias and Wilson. This discovery established the idea of a hot, singular origin for the Universe, and subsequent measurements affirmed the isotropy of the Universe to better than one part in 10^4 .

Cosmology thus became a physical rather than philosophical science, and with this change of emphasis came new questions. Why is the Universe so isotropic? Was it born that way, or did physical processes conspire to create the present isotropy? How did the present observed structure (galaxies and galaxy clusters) originate? What was the nature of the cosmic singularity?

These are the complex and often controversial issues which Derek Raine presents in this book without submerging the reader in a maze of equations and concepts from differential geometry. Of course, to go beyond a simple description of the issues requires more in the way of mathematical hardware, but that step is generally covered by adequate references to the more specialist literature. The discussions of "bang" and "whimper" singularities, and of rotating universes are particularly well done at this pedagogical level, though there are a few places where the going is quite hard (the Bianchi classification of homogeneous three spaces is perhaps the most challenging concept in the book).

Mach's principle, the "anthropic principle", the isotropy of the Universe and the arrow of time are all mentioned and put in context (but for some reason there is

no reference to Paul Davies' book *The Physics of Time Asymmetry*, published by Surrey University Press in 1974).

The Isotropic Universe is not confined to issues arising out of general relativity and the wish to explain space-time structure, though these are certainly the strongest and best-organized sections. The table of contents reads like a check-list of 95 topics every enthusiastic cosmologist should know something about, whatever his or her speciality. With a mere 250 pages in the book some topics get a rather brief treatment, though Raine succeeds in hardly ever lapsing into a "lecture-notes summary" style and maintains a consistent level of pedagogy. It is the pedagogical style that is, in my opinion, the book's strongest point. The discussion of the use of correlation functions and power spectra in studying the clustering of galaxies is a noteworthy example of this. As is often the case in teaching this subject, the mathematics becomes fairly complex for several pages. However, there are some good examples of the use of correlation functions and the two-dimensional power spectrum is treated via a projection of the sphere, thus eliminating the use of spherical harmonics.

So, who might benefit by reading this book? Although the author describes it as

Peaceful Uses of Nuclear Explosives

Lynn E. Weaver,
Editor

348 pages, \$8.50
ANS Order No.
410002



Nuclear explosives engineering education, from the technologies required to curriculum development. The theme was organized in five sections dealing with: (1) status of nuclear explosives engineering in 1969, (2) technological requirements of nuclear explosives engineering, (3) legal problems and educational programs, (4) university research and manpower needs, and (5) educational development.

Send check or money order to:

AMERICAN NUCLEAR SOCIETY
555 North Kensington Avenue
La Grange Park, IL 60525 USA
ATTN: Accounting Department

an intermediate-level introduction, the unaided physics or mathematics graduate may have difficulty in many places owing to a lack of sufficiently diversified background knowledge. The professional mathematician or physicist should experience less trouble in this regard. However, as a teaching book for a graduate course, its style and organization are excellent. My major concern in this respect

lies in what I would regard as an aberrant choice of units (consistency with other noted texts in this field would have been most welcome) and a price that virtually excludes all but the most enthusiastic students. □

Bernard Jones is at the Institute of Astronomy, Cambridge, and a Lecturer in the Department of Applied Mathematics and Theoretical Physics, University of Cambridge.

Molecular stickiness and its consequences

J.M. Creeth

Protein-Protein Interactions. Edited by C. Frieden and L.W. Nichol. Pp. 403. ISBN 0-471-04979-4. (Wiley: 1981.) £36.95, \$66.45.

THE past year has seen the publication of several quite advanced and thought-provoking texts in physical biochemistry — Spragg's *The Physical Behaviour of Macromolecules with Biological Functions*, Cantor and Schimmel's multi-volume *Biophysical Chemistry* and Richards's more physically-inclined *Introduction to Physical Properties of Large Molecules in Solution* spring immediately to mind. Now we have *Protein-Protein Interactions*, and one should say immediately that it overlaps but little with the more general texts. It is a book which will satisfy the specialist, although the general biochemical reader might find it rather hard going in parts. Its main function will probably be to provide a point of take-off for those research workers facing a suspected problem in protein interaction, and in this capacity it should be excellent (although it must be noted that gel-forming protein systems are not included).

The range of topics covers both the essential ground work of elucidating protein structure and behaviour in solution — two excellent chapters, these, with much of the mathematics compressed into an appendix — together with quite detailed surveys of the methods currently accepted as most useful. Winzor covers mass-migration (including its many recent refinements), and there is an interesting account by Cox of the successful computer simulation of sedimentation velocity. Jeffrey's chapter on equilibrium methods deals mostly with sedimentation equilibrium, as one would expect from the vast amount of work done in the past decade. Although the methods may be used to analyse associations — even if complicated by heterogeneity, non-ideality and pressure-dependence of specific volumes — often, as is made plain, the unambiguous solution is singularly hard to obtain.

Hammes describes the extremely sensitive fluorescence methods and Frieden the more specialized techniques available

for interactions involving enzymes. Timasheff discusses the topical problems inherent in the self-assembly of rod- and tube-like polymers, and a final chapter by Nichol and Winzor deals with the profound questions of how biological control mechanisms can arise through interactions among macromolecules and between them and small molecules. A measure of the practical application of what may appear to be an abstract development is the inclusion here of a section on lymphocyte activation.

Whether physical biochemistry exists (or should do so) as a separate discipline has recently, and quite rightly, been questioned. Although many of us would prefer to minimize the distinctions, conscientious readers of this book will be left in no doubt as to the vitality of the discipline, and the readiness of its practitioners to tackle problems in the main stream of biology. □

Michael Creeth is on the MRC External Staff in the Department of Medicine, University of Bristol.

Potential reference

Paul Barnes

Intermolecular Forces: Their Origin and Determination. By G.C. Maitland, M. Rigby, E.B. Smith and W.A. Wakeham. Pp.616. ISBN 0-19-856611-X. (Clarendon/Oxford University Press: 1981.) £39.50, \$69.

THE four contributors to this work have clearly striven to produce a definitive reference text on the title-subject. The authors pay due respect to the classic treatise of J.O. Hirschfelder, C.F. Curtiss and R.B. Bird (*Molecular Theory of Gases and Liquids*; Wiley, 1954) but, rightly, also point out the need for a review of developments of the past quarter century.

First, they cover the historical perspective and basic statistical thermodynamic theory, then go on to discuss the relevant experimental

techniques of molecular beam scattering, properties of gases and spectroscopy. Throughout this latter part the authors mostly keep to their main task of elucidating the intermolecular potential from the experimental data. The application to condensed phase (solid/liquid) potentials is reserved for Chapter 8, prior to the final summary (Chapter 9).

The format of the book is distinctly "reference style". There are the inevitable inhomogeneities in style and a degree of repetition, but the book is not intended to be read straight through; indeed, the authors are to be congratulated on making each section self-sufficient and, with this in mind, reading schemes are included. The extensive use of valuable appendices strengthens this approach, though even more use could have been made of them in places, for example for the lengthy quantum-mechanical derivations and scattering theory in Chapter 4 on molecular collisions.

The general approach to the subject is one of caution and commendable rigour. The authors continually try to point out the lessons to be learnt from the past (for example, the excessive use of oversimplistic potentials, neglect of pairwise non-additivity, fitting of potentials to inaccurate experimental data) and to define clearly the status of current research on potentials. The deliberate omission of hydrogen-bonding and ionic interactions is a key factor — this allows them to maintain their rigour throughout the book, though in doing so it could rob them of a larger "potential" readership.

This book will be well suited for those wishing to gain a good grounding in the use and elucidation of intermolecular potentials for the simpler systems, and also for those who may need to extract just part of that wealth of detail. Among such people will be computer simulation workers who, incidentally, may feel that the debt of the subject to computer simulation (or numerical quantum-mechanical calculations) has not been fully acknowledged; only three pages are directly devoted to computer simulation (in Chapter 8 on condensed phases), yet it can be argued that the development of computing is a primary reason for the advances made since the work of Hirschfelder, Curtiss and Bird, and also represents one direction where many further advances are to be expected. For all this, Chapters 7, 8 and 9 (on spectroscopic measurements, condensed phases, and our present understanding of intermolecular forces) are particularly enjoyable and are recommended for inclusion in a "first read".

Despite the somewhat cautious approach and the limited range of bonding species discussed, it is to be hoped that many readers will sample this worthy book. □

Paul Barnes is a Lecturer in the Department of Crystallography at Birkbeck College, University of London.

Biochemicals

The increasing cost of consumables

THE Medical Research Council of the United Kingdom, in its annual report for 1980–81 published two weeks ago, drew attention to the consequences for its support of research projects of the increasing costs of consumables. In reality, the council appears to have relatively little to complain of, for it was apparently able to persuade the Advisory Board for the Research Councils (which advises the Department of Education and Science on the division of research funds between the research councils) that some allowance should be made for the way in which the costs of consumables have increased more quickly than retail price indices in the past few years. Understandably, however, there are fears for the future, and not merely in Britain. When all grant-making agencies' budgets are stagnant (or falling), is there a danger that research programmes will be cramped for want of the funds with which to buy consumables?

The simple answer is yes; in the future as in the past, the lack of petty cash (or the authority to sign order forms) will prevent some investigators from doing what they wish. The more important question is whether the investigators that matter, those likely to break new ground, will be hampered in their work. There, most probably, the answer is no. There are good reasons why the fields in which the cost of consumables is most onerous are also novel fields, which most grant-making agencies are likely to have backed with some enthusiasm and thus generosity.

Recombinant DNA

Thus it is that recombinant DNA research has become, a little to its practitioners' surprise, one of the fields in which the cost of consumables has become most conspicuous. One unit working in the field now reckons to spend roughly a third of its total costs (salaries, but not heat, light and general overheads) on consumables. Radioactively-labelled compounds and restriction enzymes are the chief cost, with fetal calf serum (at about £100 a litre) a steady drain on the budget because the laboratory is also involved with culturing animal cells. The unit reckons that it would be spending even more on consumables if it bought in, as it might, regular supplies of dinucleotides and perhaps even custom-synthesized oligonucleotides — a milligramme of a 14-unit oligonucleotide synthesized to order will typically cost \$4,000.

Part of the explanation of these high costs is nevertheless readily understood — and widely accepted by potential customers. In this field as in others in the past, the materials that laboratories buy from suppliers include many which would,

until recently, have been made in the laboratory by the research people themselves. Synthetic oligonucleotides, used for example as probes for particular DNA structures within genes or as building blocks for the synthesis of larger pieces of genetic material, are a good illustration. Indeed, many laboratories prefer, at least for the time being, to make such materials for themselves on the grounds that, in that way, they can be sure that they have synthesized the intended product. But the skilled time required can be substantial.

Why buy?

Thus the alternative of buying in from a commercial company is in essence a means of augmenting a laboratory's manpower. Moreover, even at the prices at which materials are supplied, buying-in may be entirely economic when account is taken not merely of the salaries of those who would be occupied with in-house manufacture but of the way in which senior people would be diverted from more creative work by the need to synthesize, or supervise the synthesis of, straightforward chemicals. The old proverb "Penny wise, pound foolish" applies, even though all grant-making bodies may not yet agree.

These circumstances are neither novel nor unfamiliar. Most types of scientific instruments, for example, have usually begun their existence as custom-built machines developed by a single investigator. Even when commercial machines appear on the market, some investigators prefer to make their own versions, partly for the sake of extra sensitivity or flexibility. The point at which a laboratory decides to rely on commercial instruments must necessarily vary from one place to another. One telling consideration is usually the question of when academic laboratories can no longer match the engineering quality of commercial instruments. But there are frequent examples, perhaps more common in the physical sciences, where commercial versions of valuable hand-built equipment never appear, with the result that investigators must reconcile themselves to remaining partly instrument engineers.

One of the reasons for some of the confusion in the commercial supply of chemicals used in recombinant DNA research is linked with the rapid growth of interest in the field at laboratories not previously involved. At the same time, the widespread sense that the recombinant DNA field offers opportunities for entrepreneurs seems to have stimulated the formation of a number of small commercial organizations offering to supply a comprehensive list of materials in some narrow field — dinucleotides and

oligonucleotides for example. As things are, one man, a few technicians and an automatic nucleotide coupler may constitute a business.

Classically, the economists would predict that these circumstances would first stimulate competition, driving prices down. But it is by no means clear whether the market is sufficiently price-sensitive to have this effect. Another possibility is that the newer smaller companies will find it convenient to operate under the wings of larger and well-established concerns more skilled at distribution but anxious (as always) to supply as wide a range of biochemical materials as possible.

Signs of change are already under way in the market for monoclonal antibodies, the most immediate potential source of income for the host of genetic engineering companies set up in the past few years. Eventually, no doubt, some monoclonal antibodies will turn out to have a substantial market, perhaps in the purification of interferon or other proteins with a large potential usefulness in the pharmaceutical industry.

No doubt the market for these products will be determined internationally by means of competition between the suppliers. More specialized antibodies, used only in research, are likely always to be more highly priced. Their suppliers will, however, always be well aware that users in research laboratories can make their own — and will not be prevented from doing so by whatever patent protection applies to the commercial use of the materials. Moreover, many small suppliers must be acutely aware that their costs do not rise linearly with the quantities of these materials they can sell. Competition and sales apparently account for the quite sharp reduction of the prices of monoclonal antibodies in the past few months.

The prospect, then, is that the present high cost of the specialized biochemicals that recombinant DNA research has foisted on itself may well be abated as the years or even months go by. Their places at the top of the lists of expensive biochemicals will no doubt be taken by other materials whose importance is not yet apparent — and which may not yet exist.

No saving

Whatever happens to the cost of individual biochemicals, however, there seems no doubt that the cost of consumables in biochemically inclined laboratories will continue to increase. Several laboratory managers report that in the past few years they have been spending increasing amounts on restriction enzymes as increasing numbers of research groups take an

interest in recombinant DNA research. In real terms, the unit costs of these materials seem not to have risen in the past year or so, although they appear also not to have fallen as increased usage might have made possible.

Other natural products are a less predictable burden for laboratory accountants. The recent vagaries of the market in fetal calf serum seem to have been a particular source of trouble and for a variety of obscure reasons.

Fetal calf serum is an essential ingredient of much tissue culture work, but for reasons that remain obscure. The material itself is a by-product of abattoirs, and its supply thus depends to some extent on farmers' willingness to see their cows slaughtered while carrying calves, which is in turn a function of their estimate of the future trend of agricultural prices.

Twelve months or so ago, supplies of fetal calf serum were exceedingly hard to come by, and also expensive — at the peak of the market, the price exceeded £150 a litre. The explanation seems to have been that a fluctuation in the supply of fetal calves coincided with the decision of several commercial companies to produce substantial quantities of interferon from cell lines cultured in the presence of fetal calf serum.

Limited choice

At the height of the shortage, investigators were perhaps less distressed by the price of fetal calf serum than by suppliers' suspension of their previous practice of supplying samples from several batches of serum so as to allow the choice of a batch found suitable for intended tissue cultures. Although this practice has now been restored, and many laboratories have also found ways of economizing in their physical use of fetal serum — either by using less of it or by substituting other materials (such as newborn or even adult bovine serum) — anxiety persists that demand for material may eventually permanently overtake the natural supply. The question naturally arises whether the genetic manipulators should not now set out to help their colleagues in tissue culture laboratories by identifying the essential ingredients of fetal calf serum and, if need be, making them artificially.

The third important area of spending on consumables that seems to laboratory managers a heavier burden now than previously is the purchase of radiochemicals. As with restriction enzymes, the underlying difficulty is that laboratories are now using more of these inherently expensive materials, and in a greater variety of formulations.

The cost of radiochemicals is to a large extent determined by the salaries of those who make them. And even though there has been in the past few years a substantial growth of the demand for certain materials, small batch-production is still necessary for most radiochemicals.

Interest in recombinant DNA techniques has for example enormously stimulated the consumption of nucleotides labelled with ^{32}P . But with its short half-life of 14 days, and the usual problems of chemical degradation caused by radioactive decay, it is still necessary to supply batches of material against firm advance orders.

The manufacturers are shy of saying how much new business has been stimulated by recombinant DNA research, although some estimate that the demand for some of the more widely used radiochemicals is doubling every few months. Both the principal suppliers (Amersham International and New England Nuclear) are seeking to meet the needs of their new customers by making up radiochemicals in more convenient formulations — labelled nucleotides, for example, may be supplied in aqueous solution (thus avoiding one tedious preliminary step).

The trend towards the supply of radiochemicals with greater specific activity has also continued. Ultimately, radiochemicals of all kinds are used in some form of assay, the sensitivity of which depends inherently on the fraction of atoms at labelled site which are radioactive. Some of the ^{32}P nucleotides now available have a specific activity that is more than half the theoretical maximum, implying that more than half of the nominated phosphorus atoms consist of ^{32}P . Superficially, these materials are not disproportionately more expensive than others with less specific activity. To investigators, enhanced sensitivity and the opportunity to use smaller quantities of material are advantages that probably always outweigh the extra cost.

Even so, quite small laboratories using substantial amounts of radiochemicals consider it well worthwhile to make special deals with the suppliers. The usual arrangement is that a laboratory will undertake to buy during a year radiochemicals of more than some specified value, winning a discount on catalogue prices in return.

Although laboratories which are parties to such special arrangements are usually fiercely proud of having wrung a tough bargain from the suppliers, the arrangements are probably mutually beneficial. In the radiochemicals business, distribution costs are inevitably high, with the result that there must be substantial economies to be won from the repeated use of some well-known distribution route. But there are obvious limits on the extent to which the pooling of business from several laboratories can be used to negotiate even larger discounts from the suppliers of radiochemicals — people need to order their supplies individually and cannot risk the delays, real or imagined, that some central purchasing agency would imply.

So how much does it cost to support with consumables a scientist at the bench? This question, beloved of grant administrators, is now virtually unanswerable. One recombinant DNA laboratory in Britain

estimates that the annual cost of consumables for its particular programme now exceeds £3,500 a year for each working scientist (graduate students and technical assistants included). A similar laboratory in the United States puts the annual cost at "getting on for \$10,000". There is some evidence that larger costs may be incurred in industrial research laboratories working in these expensive fields, no doubt because those working in a commercial setting are more likely to be encouraged to augment their own labour by means of materials bought in from outside.

The relatively high cost of consumables in some fields, however, appears not yet to have much affected the rule of thumb that large laboratories with a diversified programme of research tend to spend about 10 per cent of their total budget (heat and light included) on consumables. Electron microscopists, or those working on X-ray diffraction analysis, can after all get by well enough with photographic film and perhaps a little liquid nitrogen. There is no general rule that capital equipment obviates the need for expensive consumables, as those working with radioimmunoassay know to their cost. Yet it does appear that in any large and diversified laboratory, there will be some investigators whose spending on consumables is for practical purposes negligible.

Many investigators complain, however, that the degree to which the cost of consumables must vary widely from one laboratory to another, according to the nature of the research programme, has not yet been fully appreciated by the grant-making agencies. Moreover, they argue that in the writing of research grant proposals, it is virtually impossible to calculate with any semblance of accuracy the quantities of expensive materials that will be consumed over a period of, say, three years ahead. The agencies, however, have their own internal needs of tidy accounting, although those which are the chief sources of support for research programmes in molecular biology appear to have devised effective ways of functioning flexibly.

Supplement to grants

The ideal, from many points of view, would be that grant-making agencies should agree that grant recipients in the fields in which consumable costs are highest should reconcile themselves to regular supplementary requests for funds, perhaps at intervals of six months or so during the tenure of a research grant. At least where university researchers are concerned, there is now little chance that costs can be met from university budgets.

Small research groups such as these are also less able to insulate themselves from the cost of consumables than research laboratories and institutes with their own integrated administration. In such laboratories, administrators seem to be increasingly self-conscious about the costs of consumables, and more eager than ever

to devise ways of making research groups more aware of the high costs that may be involved in their work.

Bulk purchasing seems to be one of the most commonly used devices for keeping expenditure within bounds, and appears to be most effective when applied to what may be called "dry goods" — rubber gloves and small but frequently used items of plastic ware, the cost of which seems to have increased substantially in the past few years. (The common explanation, that plastic materials are oil-based, and that their increased cost reflects the increased international cost of oil, is not all that convincing, for the quantities of hydrocarbon involved are only very small.)

Bulk contracts

In this spirit, many laboratories appear to have been able to negotiate contracts for the supply of small disposable but widely used items of equipment which offer substantial discounts, sometimes more than 30 per cent, from catalogue prices. Under these arrangements, laboratories may place an order covering twelve months ahead, drawing off supplies as needed from the manufacturer (and thus avoiding the cost of storage in the laboratory and the risk that materials will be lost, damaged or used profligately). Larger laboratories are sometimes able to buy widely used consumable items by asking suppliers to tender for them.

Despite the financial advantages there may be from such arrangements, however, research administrators seem well aware of the limitations of bulk supply. Requiring that orders for standard items of equipment should be channelled through a central office rather than directly from a laboratory to the supplier can cause delay. Moreover, the problems of identifying in advance a laboratory's need of common items of disposable equipment are in themselves formidable. The results may often be uncertain.

There are also limits to the extent to which individual investigators can be persuaded to order biochemicals from some standard supplier, able and willing to supply in bulk. People are rightly unwilling to change from one source of supply to another in the middle of a series of experiments. Equally, there may be frequent occasions when it is necessary to follow some other investigator's recipe for carrying out some experimental procedure.

In the circumstances, laboratory administrators appear to accept, the best hope for economy is that investigators themselves will be conscious of the cost of the materials they propose to use. Those whose research is supported by identifiable grants are encouraged to examine the bills their work generates, signing the invoices personally when this is appropriate. The risk that over-spending may make it necessary to look for extra funds from departmental or institute resources is usually a powerful restraint. The

competing interests of colleagues are more effective spurs to economy than pleas from accountants and managers.

In many laboratories, invoices are centrally scrutinized as a matter of routine, but usually with discretion and tact. Given the wide currency of the impression that the costs of consumable materials have suddenly become onerous, it is surprising that there are apparently few complaints of bumbledom and over-zealous interference.

One of the most effective ways of engendering a sense of economy among investigators, however, is to arrange that they are notionally required to pay even for commodities and services that would normally be provided free of charge. Such a system is operated at the Institute of Cancer Research in London, and covers the provision of services as different as electron microscopy, photography and the supply of laboratory animals.

For practical purposes, close on a fifth of the total cost of the institute's operation is charged out internally in this way. For simplicity, the cost of various units of service is counted in simple standard units — the cost of producing one electron micrograph is reckoned to be very much like that of another. The products of the animal house are reckoned in "mouse units" — the cost of a single mouse. One rat supplied is reckoned the equal of two mouse units, and so on.

Self service

Broadly speaking, each of the service departments is required more or less to break even in the course of a year, covering its own salary and equipment costs but not counting general overheads (heat and light) or the cost of management. The result is that the internal services have on the face of things a cost advantage over external suppliers, at least if they operate on a reasonably economical scale. One result has apparently been to persuade investigators to think twice before signing maintenance agreements for equipment which can technically be serviced by the institute's own electronics department.

Both the managers and the users of this system speak well of its effects, psychological though they may be. Investigators apparently welcome the chance to spend their notional money with the service departments, thus avoiding the need to wheedle and cajole people into compliance with their needs.

Such systems for sharing out the internal costs in laboratories have obvious snags, not the least of which is that the supposed arms-length relationship between the service departments and their customers must in practice be modified by a regular exchange of information between the two sides so that the suppliers can make a reasonable attempt to match their supply with the likely demand. It would serve very little purpose if an institute's animal house became a large and efficient producer of laboratory mice if the chief need had

switched from mice to rats.

There is also the obvious danger that such laboratories might become so well used to the convenience of this service arrangement that they cease to be vigilant in calculating their real costs. In the long run, the only effective safeguard is that investigators should be free to buy from outside their own laboratory if the prices there are potentially more attractive. Perhaps the most serious difficulty is that internal service departments are likely to enjoy comparatively low prestige, and thus less ready access to the funds necessary if they are to be properly equipped so as to remain efficient.

Obviously the scope for such arrangements is confined to laboratories with a certain amount of freedom to determine how resources are divided internally. The practice closely follows that in many industrial research and development laboratories, where the cost of supplies drawn from central stores will commonly be used as a management tool for assessing the cost-effectiveness of a research programme.

It is understandable enough that laboratory managers — or at least the more effective among them — should be more proud of the devices by means of which they have encouraged economy among investigators than they are distressed by the cost of consumables. At the same time, it does appear that the area of research in which the cost of consumables has recently sharply increased is probably sufficiently narrow not to have affected the general cost of research.

Objective figures are hard to come by. A document published twice a year in the United Kingdom by the Committee of Vice-Chancellors and Principals, and which purports to provide a breakdown of the general university costs, shows that the cost of a broad category which includes laboratory consumables has, in the past few years, increased at a pace almost identical with that of inflation. But this index, known as the "Brown index" after its original compiler, is probably too crude a measure to reveal the consequences of the increased cost of consumables in those fields in which the pattern of people's research has changed.

One feature of the recently discovered high cost of consumables in some hitherto shoestring fields such as molecular biology is, however, plain. While many grant-making agencies may have learned to adapt to these high costs, in present circumstances there is very little hope that university laboratories will be able to meet them out of their own departmental allowances. Moreover, the provisions now made for graduate education fall far below the costs likely to be involved. So while there is no direct evidence that the work of active and well-supported groups is at present being restricted, there is a danger that entry into the field may be hampered by the high cost. □

CORRESPONDENCE

Continued from page 302

results should the crops be attacked by more than one pest. Chemical pesticides are still the main component of most methods of integrated control and several progressive companies have now developed insecticides (chlorvinphos) that are particularly suitable for this purpose.

Integrated pest management is now accepted and promoted by WHO and FAO wherever scientific agriculture and progressive health programmes exist. No doubt wider use of this approach is desirable but this depends largely on the availability of well trained specialists and on the understanding by farmers of the need for restraint in the use of chemicals. In this respect the paper by Chapin and Wasserstrom is of some value, although its sensational presentation (with sub-titles such as "Deadly link") undermines the credibility of the authors involved.

L.J. BRUCE-CHWATT

Member of WHO Expert Panel on Malaria,
London, UK

1. WHO Official Records of the WHO 176, 106-124 (1969).
2. Akhtar, R. & Learmonth, A. *Malaria* (Research Paper No.3, Science Faculty, Open University, 1979).
3. Ray, A.P. *J. Communicable Dis.* 9, 145-171 (1977).
4. Rao, T.R. *Sisir Kumar Mitra meml Lect.* (Indian National Science Academy, New Delhi, 1978).
5. Agency for International Development *Audit Report* 76-348, 37-40 (1976).
6. WHO tech. Rep. Ser. No.549 (1974).
7. WHO tech. Rep. Ser. No.585 (1976).

SIR — It is generally agreed among malariologists that agricultural insecticides have made a contribution to selection for insecticide resistance in mosquitoes and that such resistance has made a contribution to the resurgence of malaria in Central America and South Asia. It is also generally agreed that one should be very careful before inferring a causal relationship from the discovery of a correlation between two sets of measurements.

No such care was exercised by Chapin and Wasserstrom (*Nature* 17 September, p.181-185). They infer that in El Salvador

lines but no actual data points. Each graph has a correlation coefficient marked on it of 0.96 or 0.99, but, especially in the case of the curvilinear relationships, the reader has no means of assessing how these coefficients were calculated. Their Fig. 1 is entitled "Effect of DDT on rice production in India, 1970-77", as if the correlation between these two parameters were a simple causal one. In fact, however, it is well known that other independent factors, notably the introduction of high yielding varieties, have both boosted rice yield and allowed and/or required more insecticide usage.

The basis of Chapin and Wasserstrom's curves (Figs 2 and 3) purporting to relate malaria incidence in India during 1969-77 to DDT usage and rice production, is unclear. In the figure (see below) I have plotted the data on malaria incidence issued by the Indian National Malaria Eradication Programme (NMEP) against the data shown by Chapin and Wasserstrom (Fig.4) for DDT usage. Certainly both quantities tended to rise over those years but so did other probably relevant factors such as irrigation. My graph shows a much less startling relationship than Chapin and Wasserstrom's because, according to the NMEP data, the minimum number of cases was higher, the maximum was lower and the malaria resurgence peaked in 1976 (and is reported to have continued to fall in subsequent years). No doubt the NMEP figures greatly under-report the true incidence of the disease, but there seems no reason to suppose that this under-reporting was greater

in the later years than in the earlier: Chapin and Wasserstrom give no information about any "correction factors" which they may have applied to the available data. I conclude that Chapin and Wasserstrom's graphs give a grossly misleading impression that there is a simple causal relationship between agricultural insecticides and malaria.

That the relationship is actually more complex is indicated by the following facts:

(1) Spraying of cotton crops has the, at least short term, beneficial side effect of suppressing mosquito populations.

(2) The large tonnage of insecticides used in anti-malaria spraying has certainly contributed to the selection for insecticide resistance in mosquitoes, as shown by the fact that withdrawal of this spraying has been found to lead to a levelling out or decline in the frequency of resistance genes.

(3) In the Gezira area of intensive agriculture in Sudan, resistance in *Anopheles arabiensis* is to malathion which is used in anti-malaria spraying and not to the other organophosphates used for spraying the cotton crop.

(4) Sri Lanka has very wisely banned the use of DDT and malathion in agriculture and reserved them for the anti-malaria campaign but has still had a hard struggle to contain and reverse its resurgent malaria problem.

C.F. CURTIS

Ross Institute,
London School of Hygiene and
Tropical Medicine,
London, UK

A reply —

SIR — We are sorry that Professor Bruce-Chwatt disagrees so radically with our interpretation of the facts surrounding malaria resurgence in India. In our own defence, however, we would suggest that the military conflict with Pakistan, the sharp fall in the flow of American aid, the temporary food shortages that took place there between 1973 and 1975, etc., do not explain the abundant entomological reports of *Anopheles* resistance which we cited in our article — even from such prosperous agricultural regions as Maharashtra and Gujarat. These reports, together with official accounts of the deliberations within WHO and FAO, constitute the major source of evidence upon which we have based our analysis.

Finally, like Professor Bruce-Chwatt, we appreciate the difficulties of implementing effective systems of integrated pest management in tropical areas. It was during the successful development of one such system in southern Mexico that the idea of writing this article first occurred to us.

In reply to Dr Curtis, it is unfortunate that in the process of reproducing our illustrations, many of the data points have become difficult to discern. Disregarding for the moment the 1979 figure on malaria incidence in India, however, Dr Curtis's information suggests an even higher correlation between DDT usage and the spread of disease than we have calculated. As for the question of decreased transmission, it may well be true that malaria in India peaked in 1976, but the 1977 figure he cites has been questioned by numerous specialists and in any case does not contradict

our argument about pesticide abuse.

As for the four points raised in his letter, we offer the following response:

(1) It is precisely the short-term beneficial effect of insecticides that encourages cotton growers and public health officials to apply them. What we have argued, however, is that the disadvantages of insect resistance soon come to outweigh these rather ephemeral benefits.

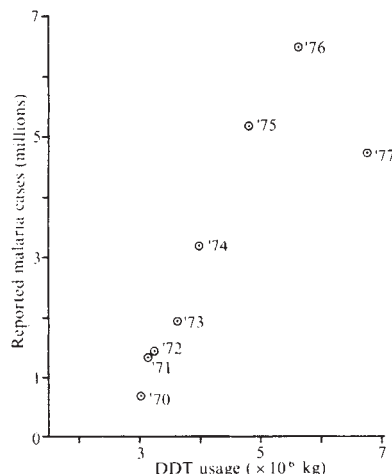
(2) It is difficult to separate the effects of insecticides used for malaria control and those used in agriculture. What is clear, however, is that the decline in resistance after anti-malaria programmes are discontinued is a limited and unfortunately rare phenomenon.

(3) Although it may be true that *Anopheles* mosquitoes in Sudan are not resistant to the organophosphates used on cotton, most countries have not been so lucky. How does Dr Curtis explain the almost complete and apparently irreversible resistance among malaria vectors in India, South-East Asia and Central America?

(4) As the case of Sri Lanka indicates, restricting the use of a particular chemical does not guarantee that mosquitoes will remain susceptible to it: application of related (and even unrelated) compounds is often sufficient to stimulate resistance. Moreover, if initial success in combating malaria leads to the reduction of screening and treatment procedures (as commonly occurs), epidemic resurgence will indeed be exceedingly difficult to control.

ROBERT WASSERSTROM
GEORGANNE CHAPIN

Columbia University,
New York, New York, USA



"each kilo of insecticide added to the environment will generate 105 new cases of malaria". Taken literally and from their own data this would imply 168 million cases of the disease in a country with a population of 4.3 million.

Chapin and Wasserstrom present three figures with apparently calculated points and